94-082
94_082-14_TJN 129...

Contract No.: 500-94-0062
MPR Reference No.: 8246-126

CMS Library
C2-07-13
7500 Security Blvd.
Baltimore, MD 21244

# Effects of Per-Episode Prospective Payment for Medicare Home Health Care on Patient Selection and Retention

## *Final Report*

***June 7, 2000***

***Author:***
***Christopher A. Trenholm***

***Programmers:***
***Dina Kirschenbaum***
***Amy Zambrowski***

Submitted to:

Health Care Financing Administration
7500 Security Boulevard (C-3-21-06)
Baltimore, MD 21244-1850

Submitted by:

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
(609) 799-3535

Project Officer:
Ann Meadow

Project Director:
Valerie Cheh

Principal Investigators:
Valerie Cheh
Randall Brown
Barbara Phillips

## ACKNOWLEDGMENTS

In preparing this report, I received support and assistance from many colleagues, for which I am grateful. Valerie Cheh and Randall Brown contributed their policy expertise and offered thoughtful comments throughout the research; Lyle Nelson provided an excellent critical review; Barbara Phillips, Valerie Cheh, and Jennifer Schore contributed to the background chapters through previous work; Cindy McClure adeptly produced the report; and Roy Grisham provided skilled editorial support with help from Patricia Ciaccio. Ann Meadow at HCFA provided very useful comments to an earlier draft of the report. Any errors or omissions in this report are my own.

# CONTENTS

Chapter Page

EXECUTIVE SUMMARY . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . xi

I THE PER-EPISODE HOME HEALTH DEMONSTRATION AND EVALUATION . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1

A. THE MEDICARE HOME HEALTH BENEFIT . . . . . . . . . . . . . . . . . . . . . . 2

B. THE PER-EPISODE DEMONSTRATION . . . . . . . . . . . . . . . . . . . . . . . . . . 4

1. Demonstration Payment and Incentives . . . . . . . . . . . . . . . . . . . . . . . . . . 6
2. Other Demonstration Procedures . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

C. GUIDE TO THE REST OF THIS REPORT AND OTHER EVALUATION REPORTS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

II DATA . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13

A. THE MAIN DATA FILE . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

1. Creating Patient Records . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
2. Available Variables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

B. SAMPLE TO STUDY SELECTION BASED ON EXPECTED SERVICE USE (VISITS) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19

1. Creation of the Analysis Sample . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
2. Outcome Variables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
3. Control Variables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26

C. SAMPLE TO STUDY SELECTION BASED ON EXPECTED PER-VISIT COSTS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29

1. Creating the Analysis Sample . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
2. Outcome Variables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
3. Control Variables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33

D. SAMPLE TO STUDY RETENTION . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34

1. Creating the Analysis Sample . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
2. Outcome Variables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35
3. Control Variables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41

**CONTENTS** *(continued)*

**Chapter** **Page**

II *(continued)* E. SUMMARY . . . . . 44

III METHODS . . . . . 45

A. ESTIMATING OVERALL IMPACTS ON PATIENT SELECTION . . . . . 45

1. Regression Models to Investigate Selection Based on Expected Service Use . . . . . 45
2. Potential Limitations of the Analysis on Use Selection . . . . . 47
3. Regression Models to Investigate Selection by Expected Per-Visit Costs . . . . . 50
4. Potential Limitations of the Analysis on Cost Selection . . . . . 52

B. ESTIMATING OVERALL IMPACTS ON PATIENT RETENTION . . . . . 53

1. Choosing the Proper Unit of Analysis . . . . . 53
2. Regression Models . . . . . 55
3. Potential Limitations of the Analysis on Patient Retention . . . . . 56

C. ESTIMATING SUBGROUP IMPACTS . . . . . 57

D. ROBUSTNESS TESTS . . . . . 59

1. Comparing Impacts Between Demonstration Years . . . . . 60
2. Sample Weights . . . . . 61
3. Outliers . . . . . 62

E. STATISTICAL PRECISION AND OTHER MEASUREMENT ISSUES . . . . . 63

1. Hypothesis Tests for the Impact Estimates . . . . . 63
2. Design Effects . . . . . 64
3. Weighting . . . . . 65

IV IMPACT OF PROSPECTIVE PAYMENT ON PATIENT SELECTION AND RETENTION . . . . . 67

A. EXPECTED EFFECTS . . . . . 67

**CONTENTS** *(continued)*

**Chapter** **Page**

IV *(continued)*

B. IMPACTS ON PATIENT SELECTION .......... 70

C. IMPACTS ON PATIENT RETENTION .......... 78

D. ROBUSTNESS .......... 81

1. Findings .......... 82
2. Overview .......... 86

V SUBGROUP IMPACTS .......... 87

A. EXPECTED IMPACTS .......... 88

B. ESTIMATED IMPACTS AMONG SUBGROUPS .......... 90

C. SUMMARY .......... 98

VI CONCLUSIONS .......... 99

A. KEY FINDINGS .......... 99

B. LIMITATIONS OF THE ANALYSIS .......... 101

C. POLICY IMPLICATIONS .......... 103

REFERENCES .......... 105

APPENDIX A: SUMMARY OF CASE-MIX ADJUSTMENT FOR PAYMENTS DURING THE DEMONSTRATION .......... 107

APPENDIX B: SUMMARY OF THE VARIABLE USED TO ACCOUNT FOR PREEXISTING DIFFERENCES IN AGENCY PRACTICE PATTERNS .......... 113

APPENDIX C: VARIABLE CONSTRUCTION FOR THE ANALYSIS OF SELECTION BASED ON EXPECTED SERVICE USE .......... 117

APPENDIX D: DETAILED SUBGROUP TABLES .......... 125

# TABLES

**Table** **Page**

II.1 SELECTED VARIABLES AVAILABLE FOR THE ANALYSIS OF PATIENT SELECTION AND RETENTION . . . . . . . . . . . . . . . . . . . . . . . . 16

II.2 DISTRIBUTION OF THE USE SELECTION SAMPLE, BY EPISODE COUNT AND DEMONSTRATION YEAR . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22

II.3 DISTRIBUTION OF USE SELECTION SAMPLE ACROSS AGENCIES, BY TREATMENT STATUS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23

II.4 OUTCOME VARIABLES TO EXAMINE USE SELECTION . . . . . . . . . . . . . . . 25

II.5 WEIGHTED MEANS, AND TESTS FOR DIFFERENCES AMONG TREATMENT AND CONTROL GROUPS, FOR THE EXPLANATORY VARIABLES IN THE USE SELECTION ANALYSIS REGRESSION MODELS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27

II.6 COST SELECTION SAMPLE: DISTRIBUTION OF ADMISSIONS (EPISODE STARTS) ON WHICH AGENCY-LEVEL VARIABLES ARE BASED, BY TREATMENT STATUS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31

II.7 OUTCOME VARIABLES TO EXAMINE COST SELECTION . . . . . . . . . . . . . . 32

II.8 DISTRIBUTION OF THE RETENTION SAMPLE ACROSS AGENCIES, BY TREATMENT STATUS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36

II.9 OUTCOME VARIABLES TO EXAMINE PATIENT RETENTION . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38

II.10 WEIGHTED MEANS, AND TESTS FOR DIFFERENCES AMONG TREATMENT AND CONTROL GROUPS, FOR THE EXPLANATORY VARIABLES IN THE RETENTION-ANALYSIS REGRESSION MODELS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39

IV.1 PERCENTAGE OF PATIENTS ADMITTED WITH SELECTED CHARACTERISTICS THAT PREDICT A SIGNIFICANT CHANGE IN SERVICE USE DURING THE AT-RISK PERIOD . . . . . . . . . . . . 71

IV.2 PREDICTED LEVEL OF HOME HEALTH UTILIZATION DURING THE AT-RISK PERIOD, BASED ON PATIENT CHARACTERISTICS OBSERVED AT ADMISSION . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 73

**TABLES** *(continued)*

**Table** **Page**

IV.3 PREDICTED LEVEL OF HOME HEALTH UTILIZATION WITHIN THE MAJOR CASE-MIX CATEGORIES . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 76

IV.4 CHANGE IN THE PROPORTION OF PATIENTS SERVED WITH POTENTIALLY HIGH PER-VISIT COSTS (FROM THE PREDEMONSTRATION TO THE DEMONSTRATION) . . . . . . . . . . . . . . . . . . 77

IV.5 CUMULATIVE RATE OF TRANSFER DURING THE YEAR AFTER ADMISSION . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79

IV.6 RATE OF TRANSFER AND OTHER AGENCY INVOLVEMENT BY ONE YEAR AFTER ADMISSION, UNDER ALTERNATIVE MEASURES . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80

IV.7 COMPARISON OF IMPACTS ON KEY PATIENT SELECTION AND RETENTION MEASURES UNDER ALTERNATIVE EMPIRICAL SPECIFICATIONS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83

IV.8 COMPARISON OF IMPACTS ON KEY PATIENT SELECTION AND RETENTION MEASURES BETWEEN YEAR 1 AND YEAR 2 OF THE DEMONSTRATION . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 84

V.1 IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY IS FOR-PROFIT OR NONPROFIT . . . . . . . . . . . . . 91

V.2 IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY IS HOSPITAL-BASED OR FREESTANDING . . . 92

V.3 IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY IS SMALL OR LARGE . . . . . . . . . . . . . . . . . . . . . 95

V.4 IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY HAS A HIGH-USE OR LOW-USE PRACTICE PATTERN . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 96

V.5 IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY IS A CHAIN MEMBER . . . . . . . . . . . . . . . . . . . . . 97

# EXECUTIVE SUMMARY

As part of its ongoing effort to study methods of providing more cost-effective care, the Health Care Financing Administration (HCFA) implemented the Per-Episode Home Health Prospective Payment Demonstration. Under the demonstration, home health agencies received a fixed, lump-sum payment for the first 120 days of each episode of care provided to Medicare beneficiaries and a predetermined rate for each visit thereafter. By allowing agencies to retain most of any surplus payments over cost, prospective payment gives agencies a financial incentive to provide home health care in a more cost-efficient manner than under traditional cost-based reimbursement.

In this report, we examine data from the first two years of the demonstration to estimate the possible effects of prospective payment on agencies' selection and retention of home health patients. Our key finding is that prospective payment (as defined by demonstration rules) *did not significantly alter the selection or retention of patients* relative to a system of cost-based reimbursement. Prospectively paid agencies admitted patients with similar levels of expected service use and with similar types of medical conditions and other characteristics highly predictive of future service needs. In addition, their rates of transfer were similar and quite low; by one year after admission, prospectively paid agencies had transferred only about three percent of their patients to another agency.

## DEMONSTRATION OVERVIEW

Ninety-one agencies in five states entered the three-year demonstration at the start of their 1996 fiscal years. Prior to the start of the demonstration, the participating agencies were randomly assigned to either the treatment or the control group. Agencies assigned to the treatment group were reimbursed under the demonstration's prospective payment method, while those assigned to the control group continued to be reimbursed under cost-based reimbursement (the payment method Medicare used for all home health agencies when the demonstration began).

For each patient admitted to home health care, treatment group agencies received a lump-sum payment for the first 120 days of home health care, regardless of the number of visits provided or their cost. (Durable medical equipment, nonroutine medical supplies, and Part B ambulatory services continued to be reimbursed at cost.) The amount of the per-episode payment was based on each agency's own costs in the fiscal year immediately preceding its entry into the demonstration, adjusted for changes in its case mix and for inflation. An agency could receive a new per-episode payment for a given Medicare beneficiary only after the 120-day "at-risk" period had ended and a 45-day gap in services had taken place. For each visit after the at-risk period that did not begin a new episode, treatment agencies received a fixed payment that varied by type of visit and was again based on the agency's own predemonstration costs.

## HYPOTHESES AND METHODS

Our basic hypothesis was that agencies receiving prospective payment would try to obtain a lower-cost patient mix in order to bring their episode costs below the per-episode payment. Two particular changes in patient selection were hypothesized. The first was that agencies would reduce their share of admissions expected to be high utilizers of care; the second was that they would reduce the share of admissions expected to have relatively high per-visit costs. With respect to patient retention, we hypothesized that prospectively paid agencies would raise their transfer rates in an effort to discharge high-cost patients more quickly.

Data for our analysis come from several sources. Medicare claims data allowed for the creation of a "main file" of patient-level records for each episode of care begun between an agency's start in the demonstration and August 1997. These data also provided extensive information on the timing and nature of patients' service use, which was used in constructing outcome variables that measured agency retention rates and patient selection patterns. Data from the demonstration's quality assurance contractor provided nurses' assessments of patients at home health admission. In conjunction with the service use data from Medicare claims, these assessments were used to create our outcome variables, which measured agency selection patterns. Additional data sources provided extensive information on agency and area characteristics, used principally to construct control variables for our regression models and define key subgroups. These sources included agency cost reports for both the predemonstration and demonstration periods, data from the Area Resource File, and selected data from the demonstration's implementation contractor.

Our analysis used three samples, each derived from the main file of patient records. These samples varied considerably in size and composition due to a combination of missing data and differences in data requirements. To the extent that we could examine whether our findings were sensitive to these missing data, they appear to be highly robust. Our first sample was used to analyze whether treatment agencies selected patients based on their expected service use. This "use selection" sample included 54,353 observations (episode starts during the demonstration) in 84 agencies. We used the second selection sample to analyze whether treatment agencies selected patients based on their expected costs per visit. This "cost selection" sample included 109,594 observations in 87 agencies. Finally, the "retention" sample, which we used to examine whether treatment agencies were more likely to transfer patients, included 86,448 observations in 79 agencies. Most of the analysis of retention focused on a subsample of 52,133 observations corresponding to admissions in demonstration year 1 because of a limited observation period (postadmission) in year 2.

Ordinary least squares models and logit models were used to estimate program effects, controlling for preexisting differences between treatment and control agencies. Observations were weighted so that each agency was represented equally in the analysis. Standard errors of impact estimates were calculated using special software designed to account for the effects of sample clustering and weighting, so as to avoid overstating the precision of the estimates. Sensitivity testing of our regression estimates showed that they were not sensitive to the weighting scheme or to outlier values. The impact estimates also displayed no change between the first and second years of the demonstration.

## FINDINGS

### No Significant Change in Access for High-Cost Patients

Per-episode prospective payment did not lead agencies to avoid patients expected to be high utilizers of care. Treatment agencies were no less likely to admit a patient with a serious medical condition, limitations in activities of daily living, or other conditions predictive of higher-than-average service needs. In addition, based on their observed characteristics at the start of an episode, the average patient in a treatment agency was predicted to use almost exactly the same level of services during the at-risk period as the average patient in a control agency (see Figure 1).

We also find no evidence that the demonstration affected the admission of patients expected to have relatively high costs per visit. For example, the demonstration had no significant effect on the proportion of patients admitted from highly rural areas or from areas with a high percentage of non-English-speaking households.

### Possibility of Slight Favorable Selection Among Readmissions

Among the roughly 1 in 10 patients with multiple episodes of care in the same agency, those readmitted to treatment agencies were found to have slightly lower predicted service use (on average) than those readmitted to control agencies. While consistent with the incentives to select a lower-cost patient mix under the demonstration, this difference appears to be due to an underestimate of predicted service use among the treatment group rather than a change in selection practices on their part. Nevertheless, we cannot reject the possibility that a small degree of favorable selection took place among readmissions to the treatment group, particularly since they are such an attractive target for these efforts.

### Transfer Rates Remained Low and Unchanged

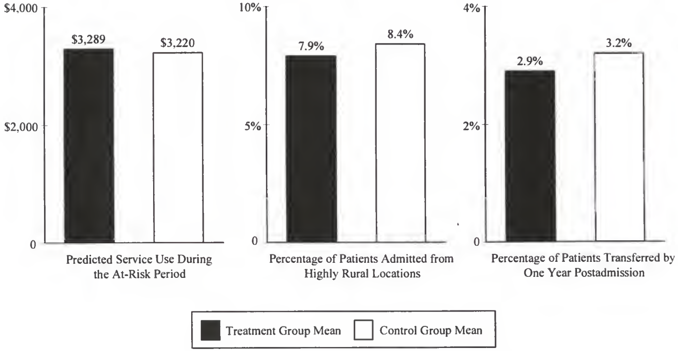
Agencies also displayed no greater propensity to transfer high-use patients after they were admitted. In each month after admission, the rates of patient switching from demonstration agencies to another demonstration agency were low and nearly identical for treatment and control agencies. By one year after admission, for example, 3.2 percent of patients had been transferred from control agencies, compared to 2.9 percent from treatment agencies (Figure 1).

### Findings Were Robust Across Different Agency Types

Impact estimates for selected agency subgroups underscore the main finding that the demonstration had no adverse effects on access to care for higher-cost patients. For the 10 subgroups investigated, most of the estimated impacts on patient selection or retention were insignificant. Moreover, among the set of significant impacts, none were consistent with the incentives to select or retain a lower-cost patient mix.

FIGURE 1

SUMMARY OF SELECTED MEASURES OF PATIENT SELECTION AND RETENTION



Note: None of the treatment-control group differences shown are significantly different from zero at the .10 significance level. The treatment group mean has been regression-adjusted.

### Some Agency Types Experienced Adverse Selection

Three types of treatment group agencies--hospital-based agencies, agencies with low-use practice patterns, and small agencies--had patients with significantly greater expected service use than their control group counterparts. After evaluating whether this adverse selection could have been intentional (or otherwise caused by the system of payment), we conclude, instead, that these agencies simply experienced an adverse patient draw that was unrelated to per-episode payment.

## LIMITATIONS OF THE ANALYSIS

Despite the significance and robustness of our findings, this study does have some important limitations. However, none of these limitations casts any doubt on the validity of the finding that agencies did not change their selection or retention practices in response to prospective payment.

Like any study design based on voluntary participation, the agencies in this demonstration may reflect a group quite different from agencies that did not participate. For example, if the agencies incentives, they might have had less reason to pursue alternative strategies to lower per-episode costs. If this is true, our demonstration impacts might understate the actual changes in selection and retention that would take place nationally under a similar payment system. However, one key factor that suggests our results can be widely generalized is that agencies of all types displayed no significant changes in their patterns of selection or retention in response to the demonstration. Although agencies of differing sizes, auspice, or practice patterns might be expected to differ, both in their capacity and their desire to respond to demonstration incentives, none of these agencies displayed an effort to admit lower-cost patients or transfer high-cost ones more quickly or more often.

A second limitation is that a national program of prospective payment would almost certainly differ from the one implemented for the demonstration. For example, agencies would not be protected from incurring financial losses under a national system, which could lead some agencies to respond more aggressively and to try to obtain a lower-cost patient mix. Moreover, an agency's case-mix-adjusted payment would almost certainly not be based on its own prior experiences, but rather on some combination of regional and national averages. For some agencies, particularly those with relatively higher costs per episode, this might lead to substantial changes in patient selection or retention as they try to bring their costs down below the payment. This behavior is more likely if the case-mix adjuster used to set payments does not adequately account for differences in expected costs.

Finally, given the strong competition in the market for home health care, many agencies probably could make only large-scale changes in their referral sources over an extended period, perhaps years in length. This suggests that the period evaluated, the first two years of a three-year demonstration, might have been too short to identify material changes in agencies' patterns of selection despite their best efforts to do so. However, the fact that we observed no significant effects on the selection or retention practices of any agency types during the period that we could evaluate suggests that important effects were unlikely in the longer term as well.

## POLICY IMPLICATIONS

The findings from this report provide key information on the effects of prospective payment at the very time when HCFA, under mandate from the 1997 Balanced Budget Act, must design and implement a nationwide prospective payment system.

### Efficacy of the Demonstration Design Further Supported

The key result from this study--that agencies did not significantly alter their selection or retention patterns in response to per-episode prospective payment--is highly relevant to the design of a national payment system. When combined with the findings of large-scale reductions in service use under the demonstration (Trenholm 2000), this result suggests that an episodic prospective payment system has the potential to generate considerable Medicare savings without causing significant access dislocations.

**Accurate Case-Mix Adjustment May Be a Key Element in Program Stability**

An additional result that is relevant to the design of a national payment system is the adverse selection that certain agencies may have experienced under the demonstration. This result underscores the need for an effective case-mix adjuster under a nationwide system of per-episode payment, since, without it, random variation in patient mix over time might place many agencies in financial trouble. For example, our subgroup results show that small treatment agencies had an expected cost per patient that was nine percent greater than the expected cost of small control group agencies. While the demonstration case-mix adjuster accounted for some of this difference (roughly one-third), this change might still have been enough to threaten the solvency of at least some of these agencies in the absence of the loss protection and other protective features of the demonstration. Moreover, an effective case-mix adjuster greatly reduces the incentive for agencies to alter their selection or retention practices to obtain a healthier patient mix. This may prove important under a nationwide program since, as noted above, an agency faced with the possibility of significant losses might be more likely to respond to these incentives than an agency in the demonstration.

# I. THE PER-EPISODE HOME HEALTH DEMONSTRATION AND EVALUATION

The Per-Episode Home Health Prospective Payment Demonstration of the Health Care Financing Administration (HCFA) tests the extent to which prospective payment for Medicare home health increases efficiency in service provision. Per-episode payment encourages efficiency by giving agencies the financial incentive to reduce their costs. Specifically, under the demonstration payment system, any savings generated from lower costs per episode of patient care might result in agency profit.[1] These incentives differ greatly from those under the current system of cost-based reimbursement, which does not provide a reward for cost containment.

In this report, we present findings from an analysis of demonstration impacts on the selection and retention of patients during agencies' first two years of participation in the demonstration.[2] These outcomes are very important to investigate because the incentive to lower costs under prospective payment might lead agencies to pursue a lower-cost mix of patients rather than (or in addition to) trying to improve their efficiency. For example, a particularly worrisome change agencies might undertake in response to prospective payment would be to reduce their admission of high-needs patients or to transfer these patients more quickly or more often. This report looks for evidence of precisely this type of behavior under the demonstration, thereby providing valuable input to the design of a payment system that can lower costs *and* ensure access to high-quality health care for all qualified beneficiaries.

---

[1] Strictly speaking, only for-profit agencies earn profits; nonprofit agencies generate surpluses that cannot be distributed to "owners" of the organization. A for-profit firm can distribute profits (or surpluses) as it wishes, including bonuses to employees, distributions to shareholders, and so forth. However, for brevity, we use the term "profits" in this report to refer both to surpluses generated by nonprofit agencies and to profits earned by for-profit agencies.

[2] Originally, this report was to focus on selection and retention during the entire three-year demonstration. However, Congress mandated that, beginning in 1999 (later changed to 2000), the Medicare program would use a prospective payment system when paying for home health services. In turn, HCFA requested that these reports be submitted before data from the third demonstration year would be available, so that HCFA could use the reports' findings in the design of the new payment system.

The rest of this chapter provides an overview of the Medicare home health benefit, the Medicare-certified home health industry, and the Per-Episode Home Health Prospective Payment Demonstration. A reader familiar with Medicare home health may wish to proceed directly to Section B, which describes the demonstration features.

## A. THE MEDICARE HOME HEALTH BENEFIT

Congress established the Medicare home health care benefit in 1965, when the original Medicare program was created. Home health benefits were included to offer beneficiaries with acute conditions a less intensive and less expensive alternative to inpatient hospital care. At different times since the inception of the Medicare program, the home health benefit has been modified, partly to increase access to care. The current benefit covers home health services mostly under Medicare Part A, and deductibles and coinsurance do not apply.[3] To be eligible for home health benefits, the beneficiary must (1) have Medicare coverage; (2) be homebound; (3) be under the care of a physician; and (4) need skilled nursing, physical therapy, or speech therapy on a part-time or intermittent basis.[4]

---

[3] In some rare instances, such as when an individual does not qualify as homebound or does not have Part A coverage, home health care may be covered under Medicare Part B. For ambulatory Part B home health, deductibles and coinsurance do apply. In addition, under the Balanced Budget Act of 1997, the long-term use of home health services was transferred to Part B. Our analysis period predates this change, however.

[4] Skilled nursing services are covered as long as (1) a physician has ordered them, (2) they are required on a part-time or intermittent basis, (3) they require the skills of a registered nurse (or of a licensed practical nurse or licensed vocational nurse under a registered nurse's supervision), and (continued...)

Coverage under the home health benefit broadened considerably after the settlement of a lawsuit against HCFA in 1989, which, in turn, contributed to dramatic growth in Medicare home health expenditures after settlement. Medicare payments for home health care rose from $2.4 billion in 1989 to $16.8 billion in 1997, more than tripling the share of total Medicare spending on home health (Health Care Financing Review 1998). Nearly all this growth has been due to an increase in the provision of services, which has coincided with a dramatic expansion in industry size. Between 1989 and 1997, the number of home health users per 1,000 beneficiaries rose from 51 to 109, and the average number of visits per user rose from 27 to 73 (Health Care Financing Review 1998; and U.S. General Accounting Office 2000).[5] Coinciding with this growth, the number of Medicare-certified home health agencies rose from 5,700 to more than 10,000 (U.S. General Accounting Office 1998). Service use patterns displayed striking differences across regions throughout this period. In 1997, for example, the average number of visits per user ranged from 37 visits in Maryland to 161 in Louisiana (U.S. General Accounting Office 2000).

The dramatic growth in Medicare home health expenditures, combined with striking regional variation in service use and recent investigations of industry fraud and abuse, prompted Congress to legislate changes to the Medicare home health benefit as part of the Balanced Budget Act of 1997. The Act mandated the implementation of a per-episode prospective payment system for Medicare home health by 1999. This was later changed to October 1, 2000 under the Omnibus Consolidated

---

[4] (...continued)
(4) they are reasonable and necessary to treat an illness or injury. Physical therapy and speech therapy are covered if a physician's assessment recommends these services. Beneficiaries who need only occupational therapy are entitled to benefits only if they have established a prior need for skilled nursing care, speech therapy, or physical therapy in the current or previous certification period (see Teplitsky and Janson 1985-1992).

[5] These figures reflect only Medicare beneficiaries in fee-for-service.

Emergency Supplemental Appropriations Act of 1999. Other changes to the home health benefit under the Balanced Budget Act of 1997 include:

- Lowering the per-visit cost limit (known as the Section 223 limits) from 112 percent of the mean to 105 percent of the median cost for freestanding agencies in the region. The Section 223 Limits were frozen for reporting periods that began between July 1, 1994, and June 30, 1996.
- Defining the maximum payment for an agency by using a new algorithm that is based on annual per-beneficiary costs or per-visit costs in a base year, whichever is lower. This algorithm is commonly referred to as the Interim Payment System (IPS), because it is intended to last only until the mandated prospective payment system takes effect.
- Eliminating coverage for blood drawing when it is the only home health service required
- Redefining "part-time" and "intermittent" care
- Redefining "service location" on the basis of the location of the patient rather than of the agency
- Requiring additional billing information (an identifier for admitting physician and visit length)

The changes in the per-visit cost limit and the algorithm for payment (the so-called IPS system) have no effect on demonstration agencies until after their participation in the demonstration has ended. However, demonstration agencies are affected by the other changes.

## B. THE PER-EPISODE DEMONSTRATION

Because a cost-reimbursed payment system provides no mechanism that enables home health agencies to earn profits by cutting costs, it offers no incentive for the agencies to provide services efficiently. Moreover, by reimbursing costs up to allowable limits, the system effectively subsidizes inefficient providers. The intent of per-episode prospective payment is to increase efficiency by using the opportunity to generate profits as the primary incentive.

Ninety-one Medicare-certified home health agencies in five states--California, Florida, Illinois, Massachusetts, and Texas--enrolled in the three-year per-episode demonstration.[6] Of these, 47 were randomly assigned to the treatment group and received per-episode payment. The remaining 44 were assigned to the control group and continued to operate under cost reimbursement. One control agency then switched to the treatment group near the start of the demonstration, leading to a revised total of 48 treatment agencies and 43 control agencies.[7] Each agency entered the demonstration and began implementing prospective payment at the start of its fiscal year. The first entrants to the demonstration began implementing prospective payment in June 1995; the latest entrants began in January 1996. Demonstration operations were planned to continue through December 1998. (The demonstration was extended for treatment agencies until prospective payment for Medicare home health care is implemented nationally, though the evaluation component ended as of December 1998.)

Mathematica Policy Research, Inc. was the contractor responsible for the process and impact evaluation of the demonstration. Several other organizations also participated in the demonstration. Abt Associates, Inc. was the implementation contractor responsible for recruiting demonstration agencies, monitoring the status of demonstration operations, and calculating certain statistics required to determine agency payments. Palmetto Government Benefits Administrator (PGBA) was the fiscal intermediary (FI) responsible for review of claims and agency payment for both treatment and control agencies. The Center for Health Policy Research (CHPR) at the University of Colorado was responsible for designing and implementing a quality assurance (QA) system for the demonstration agencies.

---

[6] Reflecting the United States more generally, considerable variation existed in the use of Medicare home health across the five demonstration states. In 1997, the average visits provided per beneficiary were as follows: California, 47; Illinois, 47; Florida, 70; Massachusetts, 89; and Texas, 134 (U.S. General Accounting Office 1998). In a companion report, we will compare the demonstration agencies with those nationwide.

[7] The switch was made at the request of an agency that had established a network with two other agencies assigned to the treatment group. The three agencies planned to merge fully at the end of the demonstration.

### 1. Demonstration Payment and Incentives

HCFA developed the Home Health Prospective Payment Demonstration to assess whether the profit motive can increase efficiency in providing Medicare home health care and thereby reduce public expenditures, without sacrificing access to care or the quality of care. Phase I of the demonstration, which tested per-visit prospective rate setting, gave agencies an opportunity to generate profits (and avoid losses) by reducing per-visit costs.[8] The evaluation of Phase I found that some agencies earned profits but did little to contain overall costs, because they had no incentive to limit the services they provided. Phase II tested a per-episode prospective payment system. This system encourages agencies to reduce their per-visit costs, and it provides the additional incentive to reduce visits per patient episode. Thus, the opportunity for cost savings was much greater under the Phase II payment system than under the Phase I system.

#### a. Payment

Agencies selected for the treatment group received a lump-sum payment for the first 120 days of home health care, regardless of the number or cost of visits provided.[9] The agencies were thus

---

[8] Although the per-visit demonstration was implemented in the same five states, most of the agencies participating in the per-episode demonstration did not participate in the per-visit demonstration. (Only agencies in the per-visit control group were eligible to do so.) For details on the results of the per-visit demonstration, see Brown et al. (1995).

[9] Durable medical equipment, nonroutine medical supplies, and Part B ambulatory home health services continued to be reimbursed at cost throughout the demonstration. In addition, if an agency did not provide one or more of the six Medicare services during the base year but began to do so

(continued...)

"at risk" for the costs of care incurred during this period. Only after the 120-day at-risk period, and after a 45-day gap in services had elapsed, could an agency receive a new per-episode payment for a given Medicare beneficiary. For each visit after the 120-day at-risk period that was not part of a new episode (referred to as "the outlier period"), treatment agencies received a fixed payment rate that varied by the type of visit. In the demonstration, a treatment agency was also paid on a per-visit basis for visits made to patients admitted before the agency began demonstration operations ("phase-in" visits), and to patients admitted within 120 days of the end of demonstration operations in that agency ("phase-out" visits).

Agencies selected for the control group received payments based on the cost-based system in place at the start of the demonstration. Specifically, the control agencies were reimbursed for their actual per-visit costs, up to 112 percent of the mean cost incurred by all agencies (for the agency's mix of visits) in the same geographic area.

### b. Prospective Rate Setting

Prospective (per-episode) rates for the at-risk period were based on a treatment agency's costs and episode profile in the fiscal year preceding its entry into the demonstration (the base year), adjusted for inflation and changes in case mix in each evaluation year.[10] The episode profile was the average number of visits provided by the agency during the first 120 days of an episode, calculated for each of the six types of visits covered by Medicare. Payments for outlier, phase-in, and phase-

---

[9](...continued)
during the demonstration, then those visits were also reimbursed at cost during the demonstration, as were the costs of care for which Medicare is a secondary payer.

[10] For more information on payment rates under the demonstration, see Phillips et al. (1995).

out visits were also based on the agency's base-year per-visit costs (adjusted for inflation).[11] HCFA's "market basket" was used to adjust both the per-visit and per-episode rates for inflation.

The case-mix adjuster classified each patient into 1 of 18 groups on the basis of 12 variables that described the patient's characteristics. From this information, an aggregate case-mix index was created for each agency. At the end of each year of the demonstration, an agency's case-mix index for that year was compared with its case-mix index for the base quarter (the last quarter of the base year). If the agency's case mix differed from its base-quarter case mix, then its aggregate payment was retrospectively adjusted (see Appendix A for additional details).

### c. Profit Sharing and Loss Sharing

To counteract the incentive to dramatically reduce services at the expense of quality, as well as prevent agencies from realizing windfall profits at public expense, HCFA shared in any profits above selected levels. If the total of a treatment agency's per-episode and per-visit prospective payments was greater than the costs for services covered by these payments, then any resulting profit greater than five percent of total allowable costs for these services was subject to profit sharing. HCFA's share was 25 percent for profits between 5 and 15 percent of total allowable costs; the share increased for profit rates greater than 15 percent by an amount that varied with the demonstration year.

HCFA provided a loss-sharing arrangement as a means of encouraging agencies to participate in the demonstration. It reimbursed treatment agencies for 99 percent of losses occurring during the first demonstration year, and for 98 and 97 percent of losses during the second and third demonstration years, respectively, as long as total payments were within the demonstration cost limits.

---

[11] Because complete data for episode profiles and settled cost reports were not available for a given year until some months after that year ended, the initial lump-sum and per-visit rates used in the demonstration were preliminary and were revised after final base-year data became available.

### 2. Other Demonstration Procedures

Three other important demonstration procedures involved the methods for billing, the process of medical review, and the requirements for QA. Treatment agencies were strongly affected by all these demonstration procedures; control agencies were principally affected by the QA requirements. In this section, we briefly discuss each procedure.

#### a. Billing

To initiate an episode of care, treatment agencies had to submit an admission bill to PGBA. Treatment agencies also had to submit interim bills for services provided during the rest of the at-risk period, although they received payments only for supplies.[12] For services provided near the start of the outlier period, an agency had to "split" its standard bill into two bills--one for services provided up to the end of the at-risk period and one for services provided from the start of the outlier period.

When a treatment agency patient was discharged during either the at-risk or the outlier period, agencies had to submit a discharge bill to terminate the episode. PGBA would not initiate a new episode for a given patient unless a previous episode had been terminated. In addition, before initiating a new episode, PGBA determined that the 120-day risk period and a 45-day gap had elapsed.

Periodic interim payments (PIPs), a means of "smoothing" cash flow for home health agencies, were originally discontinued for treatment agencies. However, a similar periodic payment system

---

[12] Agencies also had to submit interim bills to receive payment during outlier periods and to correctly calculate costs for profit or loss sharing with HCFA.

(called "biweekly interim payments") was later introduced by PGBA to address cash flow problems in some agencies.

Control agencies continued to submit bills under cost reimbursement and continued to be eligible for PIPs. PGBA based the amount of the PIP on the agency's average cost for each type of visit. For some agencies, this method differed from the practices of their predemonstration FIs, which had based the PIP on an agency's overall, average cost per visit. The new method may have led to minor changes in the PIPs that some control agencies received once they entered the demonstration.

### b. Medical Review

For agencies in the treatment group, PGBA conducted a limited medical review (known as an "abbreviated" medical review) of the care delivered during the at-risk period of the patient episode. This review attempted to determine whether the patient met the coverage criteria for home health care and whether the agency provided at least one visit that met these criteria. As a condition of payment, in addition to the admission bill, a treatment agency had to submit either (1) HCFA 485 and 486 forms (which contained information on the patient's health and eligibility status, as well as on the home health plan of treatment), or (2) clinical notes for the admission that coincided with an episode eligible for prospective payment. The medical review was based on these materials. All services paid under per-visit rate setting (for example, those delivered during the outlier period) were subject to the usual medical review, under which PGBA reviewed a sample of claims to ensure that each visit was medically reasonable and necessary.

Payment of claims depended on the outcome of the medical review. If a treatment admission claim was accepted subsequent to abbreviated medical review, the per-episode payment was provided as a lump sum. If the admission claim was denied during the medical review, processing of subsequent claims for that episode was suspended for 65 days, awaiting appeal. If an appeal was filed, these claims were further suspended until a decision on the appeal for the admission claim was made. If an admission claim was denied, but an appeal was not filed within 65 days, or if the denial of the admission claim was upheld on appeal, then subsequent claims were released for possible payment under the agency's per-visit rates.

Although all episodes initially were subject to abbreviated medical review, Medicare reduced the proportion to 25 percent in May 1996. It made this change because the claims reviews took longer than expected and because 100 percent medical review was considered unrealistic under a national program.

Medical review for control agencies continued under the nondemonstration regulations. For some control agencies, however, PGBA's review procedures may have differed from those of their predemonstration FIs, requiring them to adopt some minor procedural changes.

### c. Quality Assurance

All agencies participating in the demonstration (whether in the treatment or the control group) were required to collect and submit patient-specific information to the demonstration QA contractor. The QA procedures used a continuous quality improvement approach. Visiting staff from demonstration agencies were required to collect information (primarily on functional status and medical condition) at admission and at discharge, or 120 days after admission, whichever came first. Similar information was also collected before admission to an inpatient facility for stays of 48 hours or longer, and on a patient's return to home health care after such an inpatient stay. The QA contractor used this information to develop profiles describing patient outcomes at each agency.

These profiles were given to the demonstration agencies to help them improve the quality of care they provide.

## C. GUIDE TO THE REST OF THIS REPORT AND OTHER EVALUATION REPORTS

The next chapter of this report describes the data sources, samples, and variables used to examine patient selection and retention under the demonstration. Chapter III describes the statistical methods used to estimate demonstration impacts. In Chapter IV, we present the hypotheses and findings from an analysis of demonstration impacts on patient selection and retention. Chapter V presents demonstration impacts for specific agency subgroups. In Chapter VI, we summarize the key findings and present our conclusions.

This report is part of a series developed for the evaluation of the Per-Episode Home Health Prospective Payment Demonstration. Other reports describe the impacts of prospective payment on home health service use (Trenholm 2000) other Medicare service use (Schore 2000) and Quality of Care (Chen 2000). Future reports will include an implementation analysis and reports describing the impacts of prospective payment on home health service costs and non-Medicare service use.

# II. DATA

Our study of patient selection and retention was developed from a main data file containing 119,358 records for episodes of care beginning in the 91 demonstration home health agencies between their demonstration start and August 31, 1997. The file was built primarily from Medicare claims data extracted from the standard analytic files in May 1998. Allowing a four-month window for the standard analytic file to include all paid claims, the file contained complete Medicare service use data through December 31, 1997.[1] Because agencies entered the demonstration at the start of their 1996 fiscal years, the period covered by the file varied by agency. A majority of the agencies--57--entered the demonstration on December 31, 1995, enabling us to identify 20 months of demonstration admissions for them. The remaining 34 agencies enrolled between May 30, 1995, and November 31, 1995, for identification of 21 to 27 months of demonstration admissions.

Each record in the main file included the following information: (1) the patient's use of various Medicare-covered services from the time of home health admission (the episode start) through a maximum of 16 months, (2) a number of patient characteristics at admission, (3) the patient's Medicare service use for selected time periods prior to admission, (4) the predemonstration characteristics of the home health agency providing care, and (5) the characteristics of each agency's location.

From the main file, we identified an appropriate subset of records to form three analysis samples. We used the first sample to analyze whether treatment agencies selected patients according to personal characteristics (for example, age, health status, prognosis) associated with relatively high or low visits per episode. This "use selection" sample included 54,353 observations (episode starts during the demonstration) in 84 agencies.[2] We used the second selection sample to analyze whether treatment agencies selected patients according to characteristics (such as location) that might be associated with relatively high per-visit costs. This "cost selection" sample included 109,594 observations in 87 agencies.[3] Finally, the "retention" sample, which we used to examine whether treatment agencies were more likely to transfer patients, included a total of 86,448 observations in 79 agencies; however, our analysis focused on a subsample of 52,133 observations corresponding to admissions in demonstration year 1.[4]

---

[1] A small number of demonstration agencies experienced significant delays in billing that could have led to a substantial undercount of the number (and timing) of visits provided. This forced us to drop these agencies from the analysis of patient retention, as described in Section D of this chapter.

The remaining sections of this chapter provide a detailed description of the construction and contents of each sample. In the next section, we provide details on the main data file, which was used to form each of the three analysis samples. This is followed by three sections, each discussing the data specific to a given analysis sample; each section provides an overview of how we formed the corresponding analysis sample, a summary of the sample distribution, a description of the outcome variables, and a description of the control variables (if any) used in the empirical analysis.[5]

---

[2] Obviously, episode use and episode cost are closely correlated; indeed, our measures of use selection are weighted by the cost of each service (see Section C). Nevertheless, we prefer to define this first sample in terms of use selection, to distinguish it from the second sample that examines selection based on the per-visit cost of a patient.

[3] This analysis used an equivalent sample from a file of patient-level records reflecting episode starts in the base year immediately preceding an agency's entry into the demonstration. This predemonstration sample included 50,713 observations.

[4] As described in Section D, we focused on year 1 for this analysis because we typically could observe agencies' retention behavior for only a limited time postadmission (as little as four months).

[5] Control variables are used in multivariate regression models to control for possible preexisting differences between the patients or agencies in the treatment and control groups. For a detailed discussion of the statistical methods used to conduct our analysis, see Chapter III.

## A. THE MAIN DATA FILE

### 1. Creating Patient Records

To construct individual patient records, we used data from UB-92 bill record files obtained from the demonstration fiscal intermediary (FI), Palmetto Government Benefits Administrator (PGBA), to identify home health episodes as defined by demonstration rules.[6] We scanned the UB-92 files, beginning the file search from the date of each agency's enrollment in the demonstration, to identify the first admission for an individual and the complete set of that person's subsequent bill records. The first demonstration admission determined the episode start date from which we constructed the first record for the patient. To determine the end of the initial episode and the start of any subsequent episodes, we tracked bill records until we observed a 45-day gap in care that began after the end of the at-risk period (that is, after the first 120 days from admission). This procedure was followed regardless of whether the agency discharged and readmitted a patient within the first 165 (120 plus 45) days after the initial admission.

If we observed a readmission for a patient after 165 days, and a 45-day gap in care had taken place, we created a second record for the patient corresponding to this second demonstration episode start. We then repeated the process until we had constructed, for each patient, a series of records for all the episodes beginning in the agency between its demonstration start and August 31, 1997.

### 2. Available Variables

As Table II.1 shows, each patient record on the main file includes an extensive set of patient- and agency-level variables. These variables were used in a variety of ways, depending on the

---

[6] We used the UB-92 bill records, rather than HCFA claims data, principally because the remarks fields of these records contained additional patient-level information used in the analysis.

TABLE II.1

SELECTED VARIABLES AVAILABLE FOR THE ANALYSIS
OF PATIENT SELECTION AND RETENTION

| Patient Medicare Service Use | Patient Characteristics at Episode Start | Patient Characteristics at Episode Start | Planned Treatments for Patient at Episode Start | Agency Characteristics |
|---|---|---|---|---|
| **Source: Medicare Standard Analytic Files**<br>Selected measures of postadmission (postepisode start) visits from the admitting agency<br>Selected measures of postadmission (postepisode start) visits from non-admitting agencies<br>Calender quarter of admission<br>Length of an inpatient stay during 2 weeks before episode start<br>Whether in an SNF during 2 weeks before episode start<br>Number of home health visits in the 6 months prior to episode start<br>Total Part A Medicare reimbursement in 6 months before episode start | **Source: UB-92 Remarks Fields**<br>Has cancer<br>Has diabetes<br>Had stroke<br>Has decubiti, stage 3 or 4<br>Needs complex wound care<br>Has limitations in bathing, eating, dressing, toileting, transferring<br>Admitted to home health from hospital<br>Age<br>**Source: Medicare Enrollment File**<br>Gender<br>Race<br>Original reason for entitlement<br>Location (zip code)<br>Had Medicare for less than 6 months<br>Had Medicaid buy-in for Part A or B Medicare<br>In HMO for some time during past 6 months<br>Medicare was second payer for some time during past 6 months | **Source: QA Admission Forms**[a]<br>Health risk factors<br>Urinary tract infection<br>Incontinent or used catheter<br>Change in medical regimen<br>Clinical stability<br>Life expectancy<br>Medical prognosis<br>Functional prognosis<br>Unable to walk independently, prepare light meals, perform light housekeeping<br>Has shortness of breath<br>Is confused<br>Is depressed<br>Has emotional difficulties<br>Has cognitive difficulties<br>Requires assistance with oral medications<br>Surgical wound not healing<br>Has live-in informal assistance<br>Has paid help<br>Likely treatment can be taken over by patient, neighbor, relative, or paid help | **Source: QA Admission Forms**[a]<br>Venipuncture<br>Administration of insulin<br>Administration of other intermuscular or subcutaneous injection<br>Teach administration of insulin<br>Teach diabetic care<br>Disimpaction/follow-up enema<br>Management and evaluation of patient care plan<br>Therapeutic exercise<br>Dysphagia treatment<br>Independent/daily living training<br>Short-term therapy<br>Partial/complete bedbath<br>Personal care<br>Assist with ambulation<br>Housekeeping<br>None of the above treatments planned | **Source: Cost Reports**<br>Affiliation<br>Profit status<br>Size<br>State<br>Per-visit costs<br>**Source: Implementation Contractor**<br>Urban/rural<br>Chain member<br>**Source: Base-Quarter Case-Mix File**<br>Predemonstration practice pattern<br>Predemonstration case mix<br>**Source: Area Resource File (county level)**<br>Physicians per 10,000 persons (1994)<br>Nursing home beds per 100 elderly residents (1991)<br>Hospital occupancy rate (1993)<br>Mean Medicare reimbursement per beneficiary (1991) |

[a] These forms were completed for only about half of all episode starts.

HMO = health maintenance organization; SNF = skilled nursing facility.

analysis. Following is a brief description of the variables, organized by data source. Additional information is provided in later sections, as appropriate.

**Patient-Level Variables (first four columns of Table II.1)**

- Medicare claims data from the standard analytic files allowed us to construct numerous variables for Medicare service use following a beneficiary's episode start. In this report, we use only a small number of these postadmission variables; namely, those counting the number and timing of home health visits from admitting and nonadmitting agencies in the year after an episode start.[7]
- Medicare claims data also provided information on service use preceding the episode start. From this information, we constructed various measures of illness severity, including variables for recent acute illness (whether admitted from the hospital, length of previous hospital stay) and for home health use in the six months preceding admission.[8]
- The remarks field of the UB-92 (admission claim) provided more detailed information on patient characteristics.[9] The variables we constructed from these data included measures of impairment in activities of daily living (ADLs) and whether the patient had certain medical conditions (cancer, diabetes, decubiti) and care needs (complex wound care).

---

[7] Both Part A and Part B home health claims were used to count visits. However, the Part B visits included only those subject to a per-episode payment.

[8] We assigned the sample mean in the case of patients with incomplete data for that period and created a set of three indicator variables to identify the reason for any imputation. The indicator variables identified (1) patients in a health maintenance organization (HMO) during the six-month preadmission period, (2) Medicare as a second payer during that period, and (3) patients who were not enrolled in Medicare for a full six months. For this last group, which included predominantly beneficiaries who had just "aged" onto Medicare by turning 65 years old, we assigned the mean values for those 65.5 to 66 years old rather than the mean for the entire sample.

[9] These characteristics were used to assign patients to one of the 18-category Home Health Utilization Groups (HHUGs) that formed the basis for case-mix adjustment under the demonstration. The quality assurance (QA) admission form provided similar information but was available for only about half of all records. In the rare instance when the remarks field information was missing, we used the information from the QA form whenever possible.

- Medicare enrollment files provided basic patient demographic information at the episode start, including location (state, county, and zip), age, gender, race/ethnicity, and disability status (from the original reason for Medicare qualification).

- Data from the quality assurance (QA) admission form, collected by the Center for Health Policy Research at the University of Colorado, provided detailed information on patients' health status, care availability, and other characteristics at episode start.

### Agency-Level Variables (last column of Table II.1)

- Agency cost reports for the base year provided information on the agency's state, for-profit status, size (as measured by number of visits), and affiliation. The base-year cost report and the cost reports for each year of the demonstration also provided information on each agency's per-visit costs.

- Information from Abt Associates, Inc., the implementation contractor, allowed us to identify whether the agency was a chain or located in a rural area (as defined by the U.S. Census Bureau).

- Additional data from Abt Associates for the quarter preceding an agency's entry into the demonstration (the so-called base quarter) allowed us to create measures of each agency's case-mix distribution and practice pattern prior to the demonstration.[10]

- County-level data from the Area Resource File provided measures of physicians per population, availability of nursing home beds, and other measures of the local market.

Once each analysis sample was formed from the records on the main data file, additional variables were constructed or merged onto the sample records as needed. For example, in order to look for evidence of selection by a patient's location, we merged zip-code-level data from the U.S. Census Bureau onto individual records and aggregated these records to the agency level. Any additional variables, along with their sources, are discussed as appropriate in the sections that follow.

---

[10] For a discussion of the case-mix adjuster used in the demonstration, including the creation of variables to account for differences in case mix, see Appendix A. For a detailed description of the practice patterns variable, see Appendix B.

## B. SAMPLE TO STUDY SELECTION BASED ON EXPECTED SERVICE USE (VISITS)

One form of patient selection that treatment agencies could have engaged in was to pursue/avoid patients with individual characteristics associated with low/high expected utilization of services.

In order to investigate the possibility of such "use selection," we utilized an index of case mix, based on a multivariate model, to investigate whether the patient case mix differed between the treatment and control groups, after controlling for agency-level characteristics, area-level characteristics, and patient mix in the base year. If treatment agencies were found to have a patient mix expected to require fewer visits than control agencies (cost-weighted by visit type), then we inferred that selection behavior may have occurred in response to the demonstration. In addition, individual predictors of patient service use were examined to further explore evidence of selection behavior.

This analysis required extensive detail in order to sort patients by their anticipated home health utilization. The key source for these data was the QA admission form; as described below, however, the form was not collected for a large share of admissions, leading to a substantial loss of sample.

### 1. Creation of the Analysis Sample

We created the analysis sample through a series of exclusions from the main file. Most of these drops, roughly 85 percent, were due to missing QA information. This information was missing for all episodes that began early in the demonstration because QA forms were not collected until after May 1996, leading to the loss of about 31,000 records. In addition, most agencies did not submit QA forms for all their patients after this date, leading to the loss of another 24,000 records. The absence of QA data forced us to drop seven agencies outright. Three of these agencies left the demonstration near its start and thus had no QA information available.[11] The other four agencies were quite small, and after dropping individual admissions without QA forms, we were left with too few observations to be usable.[12]

The remaining drops were due to invalid data. The largest share reflected patients enrolled in a Medicare HMO at the time of their episode start. It would have been inappropriate to include these records because the services provided to managed care patients were not subject to per-episode payment; thus, agencies' behavior should be largely independent of the demonstration incentives. We also dropped the records for patients with Medicare as a secondary payer at the time of the episode start. Similar to those in managed care, the services provided to these patients were not subject to per-episode payment. The final set of invalid records was for patients who received no home health services during the first month after the episode start. In most cases, these records reflected "false" admissions, in which the only service provided was durable medical equipment.

The resulting analysis sample included a total of 54,353 observations (episode starts) in 84 agencies. This total includes 38 control agencies with 21,582 observations and 46 treatment agencies with 32,771 observations. Note that the smaller control agency count is partly explained by the greater share of control agencies that began the demonstration early, before January 1, 1997, leading to a particularly large number of observations excluded due to missing QA forms.

---

[11] These agencies would likely have been dropped from the analysis regardless of the availability of QA forms, since an agency participating in the demonstration for only a short period probably does not provide valid observations with which to measure demonstration impacts.

[12] As described in Chapter III, our main analysis uses sample weights that scale the importance of each observation so that each agency is represented equally. By excluding an agency with very few observations, we thus avoid placing an inappropriately large weight on a handful of potentially unrepresentative observations.

Table II.2 displays the distribution of this sample among the treatment and control groups, by episode count and demonstration year. Because of the missing QA data for the early months of the demonstration, the sample included considerably more observations from demonstration year 2 than from year 1. In addition, the sample included a small number of observations from demonstration year 3. These reflect admissions to 17 agencies that started the earliest in the demonstration (before September 1995) and thus provided the longest period over which to investigate their behavior. Most sample observations (89 percent) reflect a patient's first admission to a demonstration agency. Nearly all the remaining observations reflect the start of a second episode, while less than one percent were third or later episode starts. Only about one-third of the patients with a second episode had their first episode in the sample as well (not shown)--again, a by-product of the missing QA data for early admissions.

The analysis sample has a highly skewed distribution across agencies, which is largely a function of differences in agency size (Table II.3). The largest treatment and control agencies accounted, respectively, for about 15 percent and 9 percent of all observations in the sample, and the top three in each group accounted for roughly one-quarter of all observations. In contrast, each of the three smallest agencies in the treatment group and the control group had fewer than 50 observations, reflecting only 0.1 to 0.2 percent of their respective samples. This general skewness underscores the importance of using a weighted sample to investigate demonstration effects, since otherwise any impact estimates would be driven mostly by the behavior of just a few large agencies. For a discussion of the use of sample weights and their implications for the findings, see Chapter III.

## TABLE II.2

## DISTRIBUTION OF THE USE SELECTION SAMPLE, BY EPISODE COUNT AND DEMONSTRATION YEAR

| | Total Available Observations (Episode Starts) | | Episodes Starting in Year 1 | | Episodes Starting in Year 2 | | Episodes Starting in Year 3 | |
|---|---|---|---|---|---|---|---|---|
| | Control Group Patients | Treatment Group Patients | Control Group Patients | Treatment Group Patients | Control Group Patients | Treatment Group Patients | Control Group Patients | Treatment Group Patients |
| First Episode | 19,685 | 28,418 | 6,687 | 13,260 | 12,384 | 14,772 | 614 | 386 |
| Second Episode | 2,401 | 3,498 | 398 | 895 | 1,887 | 2,528 | 116 | 745 |
| Third or Later Episode | 149 | 202 | 1 | 0 | 132 | 180 | 16 | 22 |
| **Total Episodes** | **22,235** | **32,118** | **7,086** | **14,155** | **14,403** | **17,480** | **746** | **483** |

SOURCE: Medicare claims data.

## TABLE II.3

## DISTRIBUTION OF USE SELECTION SAMPLE ACROSS AGENCIES, BY TREATMENT STATUS

| Admissions (Episode Starts) by an Agency | Treatment Group | Percentage of All Treatment Admissions | Control Group | Percentage of All Control Admissions |
|---|---|---|---|---|
| Most | 4,776 | 14.9 | 1,998 | 9.0 |
| Second Most | 2,707 | 8.4 | 1,819 | 8.2 |
| Third Most | 2,042 | 6.4 | 1,648 | 7.4 |
| Third Fewest | 48 | 0.1 | 41 | 0.2 |
| Second Fewest | 35 | 0.1 | 39 | 0.2 |
| Fewest | 24 | 0.1 | 36 | 0.2 |
| **Total Admissions** | **32,118** | -- | **22,235** | -- |
| **Agency Average** | **714** | -- | **570** | -- |
| **Agency Median** | **427** | -- | **410** | -- |

## 2. Outcome Variables

The outcome variables for the analysis of use-selection reflected a variety of predictors of patient utilization observed at admission (Table II.4).[13] They fall into two categories. The first category, dubbed "multivariate measures," were constructed using extensive patient-level information at admission to predict the actual cost incurred by the agency during the at-risk period. The primary measure used all the patient-level variables listed previously in Table II.1 *except* the group of planned treatment variables. The secondary measure included the planned treatment variables as well.[14]

The second set of variables, dubbed univariate measures, were based on a single characteristic of a patient at admission. As described in Appendix C, each of these variables was found to have a significant impact of at least 10 percent on a patient's service use during the first 120 days after admission (relative to the mean level). Thus, each variable reflects a potentially attractive indicator by which prospectively paid agencies might try to select patients with relatively limited service needs during the at-risk period. Note that, while most variables raised expected utilization, the last two variables in the table--"paid help available," and "treatment can likely be taken over--lowered expected utilization.

---

[13] The methods used to create these variables are described in detail in Chapter III and in Appendix A.

[14] We excluded the planned treatment variables in the primary measure because they may be affected by an agency's treatment status and thus are not equivalent indicators of selection in the two groups of agencies. In particular, a treatment agency might list fewer planned treatments because it is trying to scale back on its service provision, not because of differential selection. If true, the secondary measure would tend to overstate selection changes brought about by the demonstration.

## TABLE II.4

## OUTCOME VARIABLES TO EXAMINE USE SELECTION

---

**Multivariate Measures**[a]

Predicted Home Health Utilization (First 120 Days Postadmission)

"Primary" variable derived from patient characteristics at admission, as well as measures of Medicare service use prior to admission

"Secondary" variable derived from indicators of planned treatments at admission, along with the measures used to construct the primary variable

**Univariate Measures**[b]

Has Urinary Incontinence
Has Decubiti, Stage 3 or 4
Had CVA (Cerebrovascular Accident)
Has Diabetes
Requires Complex Wound Care
Surgical Wound Not Healing
Unstable at Admission
Needs Assistance in Ambulation/Walking
Unable to Do Light Housekeeping
Life Expectancy Is More Than Six Months
Paid Help Available
Treatment Can Likely Be Taken Over (by patient, family, paid help, and so forth)

---

[a] These variables were obtained through regression models that predicted each patient's (cost-weighted) utilization of home health services under a fee-for-service payment system as a function of numerous patient-level variables observable at admission. Appendix C describes the construction of these variables in detail.

[b] These variables were estimated to change the (cost-weighted) utilization of a typical home health patient by at least 10 percent, making them attractive indicators on which to select patients. The last two variables, "paid help available" and "treatment can likely be taken over," predicted lower service use; the rest predicted higher service use. See Appendix C for further details.

### 3. Control Variables

As discussed in Chapter III, the analysis is based on regression models that control for possible preexisting differences between agencies. Table II.5 lists the control variables for these regression models and displays their sample means. The means are displayed for the entire sample, as well as separately for the treatment and control groups, in order to examine whether any substantive differences existed between the two groups.[15,16]

**Summary of Sample Means.** Roughly half the patients in the total (weighted) sample were admitted by proprietary agencies, while only 12 percent were admitted by hospital-based agencies. About one-third of patients in the total sample had been admitted by agencies that were part of a chain, and roughly one-quarter (28 percent) were in small agencies, defined as providing fewer than 30,000 visits in the base year prior to the start of the demonstration. By design, patients in the sample were admitted by agencies with an average practice pattern indexed at 1.0.[17] In addition, these agencies had, on average, equal proportions (one-fifth) of its base-year admissions in the

---

[15] The means were constructed using sample weights that gave each agency equal importance in the calculations, as this approach was the one used to estimate impacts. The significance levels for the tests of equality between the treatment and control group means did not account for design effects due to clustering because we were interested only in differences within this sample, not the population of all agencies. (This is not the case in the main analysis.) We did, however, account for the design effects associated with our use of sample weights. Chapter III fully discusses the use of weighting and clustering in our analysis.

[16] With the large sample available, sufficient statistical power exists to detect very small differences between the treatment and control groups at baseline. We therefore expected that many differences in the explanatory variables between the treatment and control agencies would be statistically significant, even when the magnitude of the difference was not materially important.

[17] Recall that the practice patterns variable accounts for preexisting differences in the number of visits provided to patients, controlling for variation in case-mix distribution. An agency with an index value above 1.0 is defined as having high-use practice patterns since, in comparison to other agencies in the sample, it provided more visits to a typical patient in the base year. Similarly, an agency with an index below 1.0 is defined as having low-use practice patterns. See Appendix B for details on how this variable was constructed.

TABLE II.5

## WEIGHTED MEANS, AND TESTS FOR DIFFERENCES AMONG TREATMENT AND CONTROL GROUPS, FOR THE EXPLANATORY VARIABLES IN THE USE SELECTION ANALYSIS REGRESSION MODELS

| Explanatory Variable | Total | Treatment Group | Control Group |
|---|---|---|---|
| **Agency Characteristics** | | | |
| Proprietary (Percentage) | 50.0 | 46.7 | 53.8** |
| Hospital-Based (Percentage) | 11.9 | 8.9 | 15.4** |
| Chain (Percentage) | 32.1 | 37.8 | 25.6*** |
| Small Agency; <30,000 Visits in Base Year (Percentage) | 26.2 | 33.3 | 17.9*** |
| Agency Practice Pattern (Index of Average Visits)[a] | 1.0 | 0.93 | 1.10*** |
| Higher-Cost Patients in Base Year[b] (Percentage) | 20.0 | 18.9 | 21.3*** |
| Medium-Cost Patients in Base Year[b] (Percentage) | 60.6 | 60.6 | 60.5 |
| Lower-Cost Patients in Base Year[b] (Percentage) | 19.4 | 20.5 | 18.2*** |
| **Area Characteristics Measures** | | | |
| California (Percentage) | 25.0 | 26.7 | 23.1*** |
| Florida (Percentage) | 8.3 | 8.9 | 7.7*** |
| Illinois (Percentage) | 17.9 | 13.3 | 23.1*** |
| Massachusetts (Percentage) | 13.1 | 17.8 | 7.7*** |
| Texas (Percentage) | 35.7 | 33.3 | 38.5*** |
| Urban (Percentage) | 84.5 | 84.4 | 84.6 |
| Physicians per 10,000 Persons (Number; 1994) | 21.9 | 21.9 | 21.8 |
| Nursing Home Beds per 100 Persons >65 Years (Number; 1991) | 5.1 | 5.1 | 5.1 |
| Hospital Occupancy Rate (Percentage; 1993) | 61.7 | 62.5 | 60.8*** |
| **Sample Size** | | | |
| **Observations** | **54,353** | **32,118** | **22,235** |
| **Agencies** | **84** | **45** | **39** |

NOTE: Individual observations (patient admissions) have been weighted to give agencies equal representation in the calculations.

[a] An index of the case-mix-adjusted average visits received by an agency's patients during the first 120 days of base-quarter episodes, relative to the average across all agencies. See Appendix B for additional information.

[b] Based on the average cost of a patient's Home Health Utilization Group (HHUG) in the base year. The higher-cost HHUGs averaged between $4,888 and $8,041 per patient (over the first 120 days after admission). The medium-cost HHUGS averaged between $2,977 and $4,728. The single lower-cost HHUG (13) averaged $2,074. For additional information on the creation of the HHUGs, see Appendix A.

*Difference in means is significantly different from zero at the .10 level, two-tailed test.
**Difference in means is significantly different from zero at the .05 level, two-tailed test.
***Difference in means is significantly different from zero at the .01 level, two-tailed test.

highest- and lowest-cost HHUGs, while the remainder were in the middle-cost categories. With respect to area characteristics, the largest share of patients in the overall sample (36 percent) had been admitted to agencies in Texas, followed by California (25 percent), Illinois (18 percent), Massachusetts (13 percent), and Florida (8 percent). Nearly 85 percent of patients were admitted by agencies located in urban areas.

Significant treatment-control differences existed for several of these agency- and area-level variables. A significantly smaller proportion of treatment group patients had been admitted by proprietary agencies or by hospital-based agencies, while a significantly higher proportion of these patients had been admitted by small agencies or agencies affiliated with a chain. In addition, the average patient in the treatment group had been admitted by an agency with a practice pattern about 15 percent below the average patient in the control group (0.93 compared to 1.10). This difference is quite striking, since it essentially indicates that the average treatment patient was in an agency that provided about 15 percent fewer visits to a typical patient before the demonstration even began. The average patient in the treatment group was also admitted to an agency with a somewhat lower-cost mix of patients in the base year. Finally, treatment group patients were significantly more likely to be admitted by agencies in Massachusetts, while they were significantly less likely to be admitted by agencies in Illinois and Texas.

These differences underscore the importance of using regression models to isolate selection effects, since they suggest that a simple comparison of unadjusted means between treatment and control group patients may misstate the effects of the demonstration. For example, the agency-level HHUG variables account for differences in the case mix of patients prior to the demonstration and suggest that treatment agencies had a lower-cost mix of patients at this time. Without these control variables, our impact estimates could misstate this preexisting difference in patient mix as an effect of prospective payment.

**Subgroup Variables.** We used five control variables to define the following pairs of agency-level categories for the subgroup analysis in Chapter V: (1) an agency's proprietary status (for-profit or nonprofit), (2) its auspice (freestanding or hospital-based), (3) its size (small or large), (4) its practice pattern (low-use or high-use), and (5) whether it was part of a chain.

## C. SAMPLE TO STUDY SELECTION BASED ON EXPECTED PER-VISIT COSTS

Similar to the incentive to avoid individual patients with high expected utilization, a treatment agency also had an incentive to avoid patients who might drive up its per-visit costs. To examine the existence of such "cost selection," we constructed a series of agency-level variables to measure how selected admission characteristics associated with per-visit costs changed between predemonstration and demonstration periods. These measures, created by aggregating patient-level information at admission, focused on two particular patterns suggestive of selection effects: (1) decreased admissions from isolated areas, and (2) decreased admissions from areas with a large non-English-speaking population.

### 1. Creating the Analysis Sample

We created our sample using patient-level records on the main (demonstration) file *and* an equivalent predemonstration file that contained all admissions to an agency over an eight-month period prior to its demonstration start. We excluded three agencies from this sample because they left the demonstration in its first few months, making it inappropriate to infer demonstration impacts from their behavior. In addition, we excluded one agency because it had only 19 admissions during the first two years of the demonstration, a number too small to look for evidence of cost selection.

For the 87 agencies that remained in the sample, only a small number of available records on the demonstration and predemonstration files were omitted from our sample. Roughly three percent of the records on both files were excluded for reasons of invalid data (described in Section B). These included patients who were enrolled in a Medicare HMO or who had Medicare as a secondary payer at the time of their admission, as well as false admissions, where no home health visits were actually provided. In addition, we lost roughly five percent of the records on both files because we could not determine the characteristics of a patient's location.[18]

The resulting demonstration and predemonstration samples included, respectively, 109,584 and 50,713 observations (episode starts) in 87 agencies. This total included 46 treatment agencies with 57,540 and 27,434 respective observations, and 38 control agencies with 52,054 and 23,279 observations. As with the use selection sample, the number of observations in these two samples varied substantially across agencies (Table II.6). For the largest agencies, for example, there were well over 1,000 observations in both the predemonstration and demonstration samples, while, for the smallest agencies, fewer than 50 observations were available in the two samples.

### 2. Outcome Variables

As noted above, the sample observations within each agency were aggregated to form a set of agency-level outcome variables measuring the change in selected admission patterns from the predemonstration to the demonstration. These outcome variables centered on two characteristics that might affect the cost of visiting a given patient (Table II.7). The first is the extent of rurality or remoteness, and the second is the extent of language isolation. Both sets of measures were derived

---

[18] As described below, we used zip-code-level data from the 1990 U.S. Census Bureau to characterize whether a patient might be particularly costly to visit. In some cases, patients could not be matched to this data because their zip code was new, or it had been omitted or entered incorrectly on the Medicare enrollment database.

## TABLE II.6

## COST SELECTION SAMPLE: DISTRIBUTION OF ADMISSIONS (EPISODE STARTS) ON WHICH AGENCY-LEVEL VARIABLES ARE BASED, BY TREATMENT STATUS

| | Treatment Group | | Control Group | |
|---|---|---|---|---|
| Admissions (Episode Starts) by an Agency | Demonstration | Predemonstration | Demonstration | Predemonstration |
| Most | 8,009 | 3,819 | 6,870 | 3,064 |
| Second Most | 3,975 | 2,648 | 3,831 | 1,528 |
| Third Most | 3,215 | 1,956 | 3,645 | 1,512 |
| Third Fewest | 52 | 42 | 84 | 40 |
| Second Fewest | 51 | 40 | 73 | 35 |
| Fewest | 35 | 29 | 67 | 32 |
| **Total Admissions** | **57,540** | **27,434** | **52,054** | **23,279** |
| **Agency Average** | **1,270** | **596** | **1,251** | **568** |
| **Agency Median** | **830** | **353** | **793** | **371** |

TABLE II.7

OUTCOME VARIABLES TO EXAMINE COST SELECTION

**Rurality Measures**[a]

Average Rate of Rurality
Proportion of Highly Rural Admissions (from zip codes that are 100 percent rural)
Proportion of Rural Admissions (from zip codes that are at least 70 percent rural)

**Language Isolation Measures**[b]

Average Rate of Area-Level Language Isolation
Proportion of Admissions from a Highly Language Isolated Area (from zip codes that are at least 30 percent language isolated)

[a] These measures were derived from 1990 zip-code-level data for individual admissions, aggregated to the agency level. The source for the data was the U.S. Census Bureau. Each measure was actually investigated twice: once defining anyone residing in a Metropolitan Statistical Area (MSA) as nonrural (that is, living in a zip code that is 0 percent rural); once ignoring their MSA designation.

[b] These measures were derived from 1990 zip-code-level data for individual admissions, aggregated to the agency level. The source for the data was the U.S. Census Bureau. A household was defined as "language isolated" if no household member could speak English well.

from 1990 zip-code-level information available from the U.S. Census Bureau that was matched to each patient's zip code at the time of admission. For example, a patient was determined to live in a highly rural location if 100 percent of the persons in his or her zip code resided in a rural location.[19] Likewise, a patient was defined as living in a highly language-segregated area if more than 30 percent of the households in his or her zip code were non-English-speaking.

Additional area characteristics that might be associated with high service costs (such as expected visit length, public safety, or community health) were not investigated because of the lack of precision in available measures. For example, to identify areas so unsafe that we might expect agencies to incur added costs to serve them, we would need crime statistics data (or very detailed proxies) at the neighborhood or census tract level. Unfortunately, such data do not exist at a national level. Nevertheless, to the extent that our findings do or do not show evidence of area selection for the outcomes we could measure, they strongly inform whether we might expect selection to have taken place in general.

### 3. Control Variables

As discussed in Chapter III, the analysis of cost selection relied on empirical models that measured the demonstration impact as the (treatment-control) difference in the change in a given outcome between the predemonstration and demonstration periods. Assuming that factors affecting cost selection (other than treatment status) remained constant over time, these models did not require the use of control variables to recover accurate impact estimates.

---

[19] We investigated the selection of patients in highly rural areas because of the potentially higher travel costs associated with visits to their home. An alternative approach would be to investigate the travel distances between the home and the agency (or its branches, to the extent they could be identified). However, we rejected this approach because the more relevant travel distance is from the home of the visiting staff, which we did not observe.

## D. SAMPLE TO STUDY RETENTION

This analysis focused on whether treatment agencies tried to cut costs by more actively transferring patients to another home health agency. The key outcome of interest in this analysis was the probability that a patient experienced a transfer from a demonstration agency during the year after admission.[20] If this probability was found to be significantly higher among patients in the treatment group than the control group, it suggests that prospectively paid agencies had altered their retention behavior in response to the incentive to lower per-patient costs.

### 1. Creating the Analysis Sample

The retention sample included 86,448 observations in 79 demonstration agencies; however, for most of our analysis, we focused on a subsample of 52,133 episodes that began in demonstration year 1.[21] This subsample included 26,384 observations in 39 treatment agencies and 25,749 observations in 40 control agencies.

Like the selection samples, the retention sample was created from patient-level records contained on the main demonstration file, following a series of exclusions. These included the removal of all admissions to 12 agencies that either left the demonstration near the start (discussed previously) or had significant delays in billing during the first demonstration year.[22] This latter group, which included eight treatment agencies and one control agency, was dropped because of concern that some of the discharge information was incorrect and thus led to undercounts in transfer rates. The remaining drops were for individual records with invalid data (described in Section B of this chapter) and/or with a prior record for the same patient. As described below, the latter drops were made because the inclusion of multiple episode starts for the same patient could lead us to misstate demonstration effects on transfer rates.

---

[20] The reason for examining transfer behavior over a fixed (that is, one year) period after admission is that it allows for unbiased estimates of demonstration impacts on retention, which is not the case if we examined transfer behavior over the patient episode. See Chapter III for a complete discussion of this issue.

[21] As detailed in Chapter III, we focused on this subsample of year 1 admissions because it allowed us to observe agencies' retention behavior over an extended period after admission (at least one full year).

[22] The process used to identify these agencies is detailed in Trenholm (2000).

The observations in the main subsample of year 1 admissions were heavily skewed across agencies (Table II.8). For example, the largest treatment agency had roughly 4,500 observations in our sample, and the largest control agency had about 3,500 observations. Alone, these two agencies account for about 17 percent and 14 percent of the sample for their respective groups. In contrast, the six smallest agencies had between 30 and 52 admissions, each no more than 0.2 percent of its respective sample.

### 2. Outcome Variables

The key outcome of interest is the likelihood of experiencing a transfer from a demonstration agency. Unfortunately, the claims data did not provide direct information on whether a transfer took place, forcing us to impute our measures for patient transfer. Ideally, this imputation would be based on the gap observed between the discharge date from the admitting demonstration agency and the date of admission to a second agency. (The shorter this gap, the more likely that a patient has been transferred). However, the claims data also did not provide the exact date of discharge so the end date on the final claim was used as the best approximation.[23] Our actual measures of patient

---

[23] During the time period covered by this report, agencies were not required to itemize their dates of service, and the best available measure of the discharge date was the end date of the final claim. In many cases, this approximation was exact because the final claim covered only a one-day period or had an end date equal to the date of discharge. However, because some agencies billed on a periodic (for example, weekly) basis, the actual date of discharge for some patients preceded our approximated date.

## TABLE II.8

## DISTRIBUTION OF THE RETENTION SAMPLE ACROSS AGENCIES, BY TREATMENT STATUS

| Admissions (Episode Starts) by an Agency[a] | Treatment Group | Percentage of All Treatment Admissions | Control Group | Percentage of All Control Admissions |
|---|---|---|---|---|
| Most | 4,497 | 17.0 | 3,508 | 13.6 |
| Second Most | 2,440 | 9.2 | 1,711 | 6.6 |
| Third Most | 1,807 | 6.8 | 1,594 | 6.2 |
| Third Fewest | 49 | 0.2 | 52 | 0.2 |
| Second Fewest | 37 | 0.1 | 46 | 0.2 |
| Fewest | 30 | 0.1 | 46 | 0.2 |
| **Total Admissions** | **26,384** | -- | **25,749** | — |
| **Agency Average** | **677** | -- | **644** | — |
| **Agency Median** | **353** | -- | **473** | — |

[a] The counts shown include only episodes starting in the first year of the demonstration, as this subsample was the main focus of our analysis. The counts for the full sample were larger among both treatment and control agencies (N = 42,042 and 44,406, respectively), but their distribution was similar. (Episodes starting later in the demonstration are examined only on a limited basis because the time period observable after admission was often very short; see Chapter III for additional details.)

transfer are thus based on the gap observed between this approximated discharge date and the date of admission to a second agency.

The specific variables for patient transfer, listed in Table II.9, are as follows. The base measure assumed that a transfer had taken place if there was a period of less than four days between the (approximated) discharge date and an other-agency admission; or, if there were fewer than eight days between any other-agency claim starting before the discharge date and a second claim starting after the discharge date. This latter definition was necessary to account for cases where the actual discharge date may have preceded the date approximated. A second, less restrictive measure lets these respective gaps be one week and two weeks, and a third, least restrictive measure lets the gaps be two weeks and one month. While the cutoff date for each of these variables is fairly arbitrary, taken together, the cutoff dates provide a strong set of measures with which to examine whether any changes in transfer took place as a result of the demonstration.

An additional variable accounts for other types of nonadmitting agency involvement in the episode of care. Essentially, these are instances where a claim from a second agency appears prior to the approximate date of discharge from the admitting agency. As noted, these instances might have occurred because our approximate discharge date was later than the actual date; however, other explanations exist that are associated with demonstration incentives. Most notably, these instances might have resulted from subcontracted services not being properly billed through the demonstration agency, or from the sharing of patient by two different agencies. Since both these behaviors would have lowered the per-patient costs of agencies in the treatment group, it was important to investigate the effect of the demonstration on this outcome variable.

## TABLE II.9

## OUTCOME VARIABLES TO EXAMINE PATIENT RETENTION

---

**Probability of Transfer**[a]

Base measure (very likely transfer): transfer assumed if there were 0 to 3 days between discharge and an other-agency admission, or 0 to 7 days between an other-agency claim starting before the discharge date and a second claim starting after the discharge date.

Less restrictive measure: transfer assumed if there were 0 to 7 days between discharge and an other-agency admission, or 0 to 14 days between an other-agency claim starting before the discharge date and a second claim starting after the discharge date

Least restrictive measure: transfer assumed if there were 0 to14 days between discharge and an other-agency admission, or 0 to 31 days between an other-agency claim starting before the discharge date and a second claim starting after the discharge date

**Probability of Other-Agency Involvement**[a]

Other-agency involvement: assumed if there was a claim from a second agency with a start date before the date of discharge

---

[a] Measured cumulatively from first admission (episode start) through one year, by selected lengths of time.

## TABLE II.10

## WEIGHTED MEANS, AND TESTS FOR DIFFERENCES AMONG TREATMENT AND CONTROL GROUPS, FOR THE EXPLANATORY VARIABLES IN THE RETENTION-ANALYSIS REGRESSION MODELS

| Explanatory Variable | Total | Treatment Group | Control Group |
|---|---|---|---|
| **Demographic Measures (Percentage)** | | | |
| Age Was Original Reason for Medicare Eligibility | 82.5 | 83.9 | 81.2*** |
| Age: <65 Years | 8.3 | 7.5 | 9.0*** |
| Age: 65 to 74 Years | 29.3 | 28.5 | 30.1** |
| Age: 75 to 84 Years | 39.9 | 40.2 | 39.6 |
| Age: >84 Years | 22.6 | 23.9 | 21.3*** |
| Race/Ethnicity: White | 81.5 | 81.7 | 81.2 |
| Female | 64.4 | 64.7 | 64.1 |
| Enrolled in Medicaid | 25.3 | 26.4 | 24.1*** |
| **Medical Conditions and Care (Percentage)** | | | |
| Cancer | 13.1 | 13.2 | 13.0 |
| Diabetes | 21.4 | 21.3 | 21.5 |
| Cerebrovascular Accident (Stroke) | 14.5 | 14.6 | 14.5 |
| Decubitus Ulcer: Stage 3 or 4 | 4.2 | 4.6 | 3.8** |
| Needs Complex Wound Care | 6.7 | 6.7 | 6.6 |
| Any Medical Condition (Among the Five Above) | 46.1 | 46.4 | 45.8 |
| **Limitations in ADLs (Percentage)** | | | |
| Bathing | 72.8 | 72.9 | 72.8 |
| Eating | 28.7 | 28.0 | 29.3* |
| Dressing | 62.8 | 61.4 | 64.2*** |
| Toileting | 39.1 | 38.0 | 40.3*** |
| Transferring | 51.0 | 50.4 | 51.6 |
| Any ADL | 78.0 | 78.2 | 77.8 |
| **Patient Prior Service Use Measures** | | | |
| Admitted to Home Health from Hospital (Percentage) | 36.4 | 34.1 | 38.7*** |
| Length of Hospital Stay Ending Within 14 Days Before Home Health Admission (Number of Days) | 3.9 | 3.6 | 4.2*** |
| Had SNF Stay Ending Within 14 Days Before Admission (Percentage) | 16.7 | 17.9 | 15.4*** |
| Total Part A (Plus Part B Inpatient/SNF/Home Health/Hospice) Reimbursement in Six Months Before Admission (Dollars) | 11,222 | 11,278 | 11,167 |
| Part A Home Health Visits in Six Months Before Admission (Number) | 12.2 | 11.9 | 12.5 |

TABLE II.10 *(continued)*

| Explanatory Variable | Total | Treatment Group | Control Group |
|---|---|---|---|
| **Other Patient-Level Variables (Percentage)** | | | |
| Had Medicare for Less than Six Months Before Admission | 1.4 | 1.4 | 1.4 |
| Enrolled in HMO in Six Months Before Admission | 0.9 | 0.9 | 0.9 |
| Had Medicare as Second Payer in Six Months Before Admission | 0.3 | 0.2 | 0.3 |
| Admission Date: 5/30/95 Through 9/30/95 | 5.6 | 4.1 | 7.1*** |
| Admission Date: 10/1/95 Through 12/31/95 | 10.2 | 8.9 | 11.5*** |
| Admission Date: 1/1/96 Through 3/31/96 | 28.2 | 27.4 | 29.0** |
| Admission Date: 4/1/96 Through 6/30/96 | 24.7 | 26.0 | 23.4*** |
| Admission Date: 7/1/96 Through 9/30/96 | 18.2 | 19.0 | 17.4** |
| Admission Date: 10/1/96 Through 12/31/96 | 13.1 | 14.6 | 11.6*** |
| **Agency Characteristics Measures (Percentage)** | | | |
| Proprietary | 50.6 | 48.7 | 52.5*** |
| Hospital-Based | 11.4 | 7.7 | 15.0*** |
| Chain | 35.4 | 43.6 | 27.5*** |
| Small Agency (<30,000 Visits in Base Year) | 27.8 | 35.9 | 20.0*** |
| Agency Practice Pattern: Index of Average Visits[a] | 1.03 | 0.95 | 1.11*** |
| **Area Characteristics Measures** | | | |
| California (Percentage) | 22.8 | 23.1 | 22.5 |
| Florida (Percentage) | 7.6 | 7.7 | 7.5 |
| Illinois (Percentage) | 19.0 | 15.4 | 22.5*** |
| Massachusetts (Percentage) | 12.7 | 17.9 | 7.5*** |
| Texas (Percentage) | 38.0 | 35.9 | 40.0*** |
| Urban (Percentage) | 83.5 | 82.1 | 85.0*** |
| Physicians per 10,000 Persons (Number; 1994) | 21.5 | 21.6 | 21.5 |
| Nursing Home Beds per 100 Persons >65 Years (Number; 1991) | 5.2 | 5.2 | 5.2 |
| Hospital Occupancy Rate (Percentage; 1993) | 61.5 | 62.2 | 60.8*** |
| **Sample Size** | | | |
| **Patients** | **52,133** | **26,384** | **25,749** |
| **Agencies** | **79** | **39** | **40** |

NOTE: The statistics presented in the table are based on episodes starting in the first year of the demonstration, as this subsample is the main focus of our analysis. With the exception of the binary variables for date of admission, the statistics for the full sample (N = 86,448) are quite similar. (Episodes starting later in the demonstration are examined on only a limited basis because the time period observable after admission is often very limited; see Chapter III for additional details.) Individual observations (patient admissions) have been weighted to give agencies equal representation in the calculations.

[a] An index of the case-mix-adjusted average visits received by an agency's patients during the first 120 days of base-quarter episodes, relative to the average across all agencies. See Appendix B for additional information.

ADL = activity of daily living; HMO = health maintenance organization; SNF = skilled nursing facility.

*Difference in means is significantly different from zero at the .10 level, two-tailed test.
**Difference in means is significantly different from zero at the .05 level, two-tailed test.
***Difference in means is significantly different from zero at the .01 level, two-tailed test.

### 3. Control Variables

The control variables for our regression models included an array of key patient- and agency-level characteristics available from our main data file. (Details on these regression models, and related methodological issues, are addressed in Chapter III.) In Table II.10, we display the sample means of these variables for patients admitted during the first demonstration year.[24] The means are presented both for the entire sample and separately for the treatment and control groups, in order to examine whether the two groups differed in any important way.[25,26]

**Patient-Level Variable Means.** As Column 1 of Table II.10 shows, more than 80 percent of the patients in the sample initially became eligible for the Medicare program because of their age, and more than 90 percent were older than 65. The largest group, roughly 40 percent, was between the ages of 75 and 84. About 82 percent of the sample was white, 64 percent female, and 25 percent jointly enrolled in the Medicaid program. At admission, about one in seven (13 percent) of the patients in the sample had cancer, and a slightly higher percentage (15 percent) were recovering from a stroke. About one in five patients (21 percent) had diabetes. A very small percentage of patients had a serious decubitus ulcer (four percent) or required care for a complex wound (seven percent).

---

[24] As noted, data limitations required us to restrict the sample for much of our analysis to patients admitted during the first demonstration year.

[25] The means were constructed using sample weights that gave each agency equal importance in the calculations, as this approach was the one used to estimate impacts. The significance levels for the tests of equality between the treatment and control group means did not account for design effects due to clustering because we were interested only in differences within this sample, not the population of all agencies. (This is not the case in the main analysis.) We did, however, account for the design effects associated with our use of sample weights. Chapter III fully discusses the use of weighting and clustering in our analysis.

[26] With the large sample available, sufficient statistical power exists to detect very small differences between the treatment and control groups at baseline. We therefore expected that many differences in the explanatory variables between the treatment and control agencies would be statistically significant, even when the magnitude of the difference was not materially important.

A large majority of patients had one or more limitations in ADLs for which we have complete measures. The most common limitations were in bathing (73 percent) and dressing (63 percent); the least common was in eating (29 percent). More than one-third (36 percent) of patients had been hospitalized during the two weeks preceding admission, and another 17 percent had been in a skilled nursing facility (SNF).

The treatment-control differences in the patient-level control variables were generally quite small, often insignificant, and displayed little consistent pattern. Relative to control agencies, treatment agencies served a somewhat higher proportion of patients older than age 85, enrolled in Medicaid through a state buy-in program, or who became eligible for Medicare because of their age. Treatment group patients had significantly higher rate of decubitus ulcer, but lower rates of limitation for three of the five ADL tasks listed (dressing, toileting, and transferring). All these differences were less than three percentage points. Finally, treatment agencies had a lower share of patients from the hospital but a larger share of patients leaving an SNF. Both differences may be explained by the larger share of hospital-based agencies in the control group, since (not surprisingly) these agencies admitted a disproportionate number of their patients from the hospital.

**Agency-Level Variable Means.** Not surprisingly, the (weighted) means for the agency- and area-level variables closely parallel those presented in Table II.5 for the use selection sample. Roughly half the patients in our sample were in proprietary agencies and, by design, about half of this group were in agencies with high-use practice patterns (as indicated by an index value just over 1.0). Only 11 percent of patients were in hospital-based agencies. About one-quarter (28 percent) of patients were in small agencies, defined as providing fewer than 30,000 visits during the year preceding the demonstration. The largest share of patients (nearly 38 percent) were in Texas-based agencies, followed by California (23 percent), Illinois (19 percent), Massachusetts (13 percent), and Florida (8 percent). More than 80 percent of patients were in agencies located in urban areas.

In contrast to the patient-level variables, there were substantial treatment-control differences in the agency- and area-level variables. Perhaps most importantly, there were significant preexisting differences in agency practice patterns. This difference--0.95 for the treatment group, compared to 1.11 the control group--essentially indicates that the average treatment patient was in an agency that provided about 14 percent fewer visits to a typical patient before the demonstration even began. Agency characteristics also differed significantly in other key dimensions, and most of these differences were materially large. Treatment group patients were only half as likely to be in a hospital-based agency, for example, but far more likely to by in a relatively small agency or an agency affiliated with a chain. In addition, treatment group patients were much more likely to be in a Massachusetts-based agency, while they were far less likely to be in an agency in Illinois.

These differences underscore the importance of using regression models to estimate program impacts, as these models control for effects on service use that might otherwise be falsely attributed to prospective payment. In particular, if the relatively low-use practice patterns evident in the treatment group were associated with more frequent transfer, then a simple comparison of mean transfer rates under the demonstration might have substantially overstated the program impact.

**Subgroup Variables.** As with the analysis of use selection, selected control variables were used to define the following pairs of agency-level subgroups: (1) an agency's proprietary status (for-profit or nonprofit), (2) its auspice (freestanding or hospital-based), (3) its size (small or large), (4) its practice pattern (low-use or high-use), and (5) whether it was part of a chain. In Chapter V, we evaluate whether the demonstration affected retention differently among these various subgroups.

## E. SUMMARY

The three samples described above provide a comprehensive set of data with which to investigate the effects of the demonstration on patient selection and retention. Nevertheless, several features of these samples, including the skewed distributions across agencies, the loss of selected observations, and significant preexisting differences between agencies in the treatment and control groups, underscore the importance of conducting a careful empirical analysis. In the next chapter, we describe the empirical models used to isolate demonstration impacts and address a number of other statistical issues involved in estimating impacts accurately. In addition, we discuss the sensitivity tests that we conducted to test the robustness of our main results and describe the specification of a model to examine subgroup effects.

# III. METHODS

Throughout our analysis of patient selection and retention, we measured demonstration impacts through regression models because they controlled for preexisting differences between treatment and control agencies that might have existed despite random assignment and because they improved statistical precision. As detailed below, these models were based on a variety of statistical techniques and units of measurement. Our primary models used weighted data that gave each agency equal representation in the analysis; however, as a test of robustness, we also investigated impacts without sample weights. All analyses used standard errors that took into account the effects of sample clustering and weighting, as described in Section D.

## A. ESTIMATING OVERALL IMPACTS ON PATIENT SELECTION

Two distinct forms of patient selection might have taken place under the demonstration: (1) selection based on a patient's expected utilization ("use selection"), and (2) selection based on a patient's expected per-visit costs ("cost selection"). Both analyses examined whether treatment-control differences existed in the observed characteristics of patients admitted into an episode of home health. Thus, the unit of measurement throughout this analysis was the patient episode, or more specifically, the patient episode evaluated at the time of admission.

### 1. Regression Models to Investigate Selection Based on Expected Service Use

We employed two basic regression models to investigate the overall impact of per-episode payment on use selection. When the outcome variable was continuous (for example, the expected utilization of a patient admitted into home health), we used the following linear regression model:

(1) $Y = \alpha + X\beta + \delta T + \epsilon,$

where:

$Y$ = a continuous episode-level outcome variable, measured at a patient's admission

$X$ = a matrix of agency-level control variables

$T$ = a binary variable for treatment status that equals one for admissions to treatment agencies and zero for admissions to control agencies[1]

$\alpha$ = the intercept term

$\beta$ = the vector of regression coefficients on the control variables

$\delta$ = a regression coefficient on the variable for treatment status

$\epsilon$ = a random disturbance term assumed to have a mean of zero (conditional on the regressors in the model) that reflects all the unobserved factors affecting the outcome variable

The coefficient $\delta$ on the variable $T$ measures the effect of the demonstration payment method on the (continuous) outcome of interest and was tested to determine whether it is significantly different from zero.[2]

When the outcome variable was binary (for example, whether a patient was diagnosed as unstable at admission), the logit model was used to estimate demonstration impacts. The structure of the logit model is as follows:

(2) $Probability\ (Y = 1) = \dfrac{1}{1 + e^{(\alpha + X\beta + \delta T)}},$

---

[1] The omitted (reference) binary variable equals one for admissions to control agencies and zero for admissions to treatment agencies.

[2] See Section E for a detailed discussion of hypothesis testing.

where $Y$ is the (binary) outcome variable and the remaining variables and parameters are defined as in equation (1).[3] Given the nonlinearity of the logit model, the estimated impact of the demonstration was not measured directly by the coefficient $\delta$ on the variable for treatment status. Instead, to estimate the demonstration impact on the probability that $Y = 1$, we used the coefficient estimates from the model to generate two predicted probabilities for each observation: one assuming that the observation belonged to the treatment group ($T = 1$), and one assuming that it belonged to the control group ($T = 0$). The impact estimate was given by the average difference between these estimated probabilities. Because the statistical significance of $\delta$ determines whether the odds that $Y = 1$ are significantly different for the treatment and control groups, we used the *p*-value on this parameter to test our hypotheses about differences between the two groups (see Section E in this chapter).

## 2. Potential Limitations of the Analysis on Use Selection

As described previously in Chapter II, a key pair of outcome variables were derived from multivariate models that predicted the level of home health use of each patient in the absence of the demonstration. The covariates in these models included a series of indicators of patient severity and other patient characteristics at admission, which were predictive of patients' level of need for home health services. (Estimation of these multivariate models is detailed in Appendix C.) The variables derived from these models thus served as effective index measures to capture the patient mix in each agency.

---

[3] For ease of exposition, all later specifications in which the outcome variable may be continuous or discrete are shown for the continuous outcomes only. In all cases, however, we use both ordinary least squares and logit models, as appropriate.

One of these two index measures was derived using all available information on the patient characteristics at admission, while the second index measure excluded information on the treatments planned for the patient. Due to concern that the planned treatment indicators might understate the relative severity of patients in the treatment group, we used the latter of these two variables as our "primary" index variable in the analysis.[4] The former variable (that included planned treatments) then served as a useful "secondary" measure, essentially providing an upper bound estimate of whether and how much favorable selection might have resulted from the demonstration.

While they were very useful constructs to explore selection behavior, both index measures were potentially subject to bias from at least three sources. First, because most of the information on patients' admission characteristics was based on reporting by the visiting clinician at the admitting agency, differences across agencies might have arisen in these outcome variables due simply to variability in reporting. If such variability was, in turn, associated with an agency's treatment status, any impact estimates based on these measures would have been subject to bias. (Indeed, it was precisely this concern that led us to drop the planned treatment indicators from our primary index measure.) The direction of this potential bias is largely uncertain. On the one hand, the incentives of the demonstration might have led the admitting clinician to be more favorable about several patient indicators, such as those associated with a patient's prognosis or likelihood for independence. If true, these measures, like the planned treatment indicators, might have made the patient mix in treatment agencies appear less severe, biasing any impact findings in the direction of favorable selection. On the other hand, the incentive also existed for treatment group agencies to make their patients look more severe at admission, either to place them in higher-cost case-mix groups or to improve their quality-assurance (QA) ratings.[5] While it seems quite unlikely that such behavior would have been widespread among clinicians, this form of gaming would have had the opposite effect, essentially biasing any impact estimates in the direction of finding no favorable selection by treatment agencies.

---

[4] The concern with including these planned treatment measures is that they could have been lower among treatment group agencies than control group agencies simply because of the incentives to cut back on service use under per-episode payment. If true, this would have tended to understate the severity of patients in treatment group agencies, introducing bias in the direction of showing favorable selection by treatment group agencies.

A second potential source of bias centered on the prior home health use of patients who had been readmitted to a demonstration agency. For this group, any treatment-control differences in past service provision might have affected their observed characteristics at the point of readmission, possibly introducing differences in patient mix that would be unrelated to selection behavior. In particular, findings from Trenholm (2000) showed that treatment patients received significantly fewer visits on average than control patients, which (all else equal) would have led patients in the treatment group to have fewer home health visits and lower Medicare reimbursement prior to readmission. Since these lagged indicators are positively correlated with anticipated service needs, such differences would have tended to bias our index measures in the direction of showing favorable selection by treatment group agencies. Alternatively, differences in prior service provision might have directly affected the severity of patients at readmission; most notably, by making treatment group patients worse off at the start of a readmission. If true, this might have counteracted any efforts by treatment agencies to readmit a favorable mix of patients, making it less likely that we would observe impacts on selection for this group.

A third source of bias might have resulted simply from omitted variables in the regression models. Most notably, if treatment-control differences existed in local market factors such as the level of competition or the service needs of available patients, it might have introduced differences in patient case-mix without any behavioral response on the part of treatment agencies. Fortunately, this source of bias is probably a second-order concern, since the combination of random assignment and regression modeling (with several control variables associated with area characteristics) should have largely mitigated this potential form of bias.

---

[5] See Appendix A for a summary of case-mix adjustment under the demonstration.

Given the concerns associated with these index measures, it was critical that we explore an additional set of outcome variables and thoroughly investigate the sensitivity of our findings. As described in Chapter II, the other outcome variables that we explored were a series of 12 indicator variables found to be strongly correlated with future service use. (These variables were identified from the regression models used to construct our index measures; see Appendix C for additional information.) While each of these "univariate" measures might also have been subject to bias from some of the same sources described above, taken in total, they provided a strong additional test of selection effects.

The formal robustness tests that we conducted are described in Section D. In addition to these formal tests, a number of other sensitivity tests were conducted and summarized (though not reported) at appropriate points in Chapters IV and V. (For example, we reestimated impacts for those readmitted to home health care, excluding from the index measures all variables closely linked to prior home health use.) In total, these sensitivity tests strongly supported the basic finding that the per-episode prospective payment did not lead to selection on the basis of patients' anticipated service use.

### 3. Regression Models to Investigate Selection by Expected Per-Visit Costs

For the analysis of cost selection, we computed demonstration impacts through difference-in-difference (DD) regression models. For a given outcome, these models calculated the demonstration impact as the change at the agency level in the treatment-control difference between the predemonstration and demonstration periods. For example, assume that the average change from predemonstration to demonstration in the proportion of patients admitted from remote locations was -4.5 percent for treatment agencies and 0.3 percent for control agencies. Since, presumably, the only factor that changed differently between treatment and control agencies over this period was their form of payment, the treatment-control difference in the average change, -4.8 percent, provides an unbiased estimate of the demonstration's impact on the proportion of patients admitted from remote areas. The key benefit of the DD approach was that it controlled for all fixed differences between treatment and control agencies that might have existed prior to the demonstration. This was particularly important to this analysis since it was quite possible that significant treatment-control differences existed in agencies' service areas, in spite of random assignment.

In practice, the DD approach amounted to a simple regression model in which the dependent variable was the *agency-level* change in the agency mean for a given outcome (such as the rate of highly rural patients) between the predemonstration and demonstration, and the only regressors were a constant term and the agency's treatment status.[6]

$$(3) \quad \Delta Y = \alpha + \delta T + \epsilon,$$

where:

$\Delta Y$ = the change in a continuous agency-level outcome variable, measured by aggregating episode-level data at admission within each agency

$T$ = a binary variable for treatment status that equals one for admissions to treatment agencies and zero for admissions to control agencies

$\alpha$ = the intercept term

---

[6] All other agency characteristics are essentially fixed over time and therefore are assumed to have no effect on the change in agency outcomes.

$\delta$ = a regression coefficient on the variable for treatment status

$\epsilon$ = a random disturbance term assumed to have a mean of zero (conditional on the regressors in the model) that reflects all the unobserved factors affecting the change in the outcome variable

As in previous regression models, the key parameter in equation (3) was the coefficient $\delta$, which provided an unbiased estimate of the average demonstration impact (among treatment agencies) on the outcome of interest.

### 4. Potential Limitations of the Analysis on Cost Selection

The key limitation in this analysis was the quality of the data. As noted in Chapter II, the lack of available data did not allow us to explore impacts on several important outcomes associated with per-visits costs, including the expected visit length, the need for security, and the level of community health. In addition, the set of outcomes that we did focus on (associated with rurality and language isolation) could not be linked to the individual patients and instead had to be inferred based on the extent language isolation in the patient's zip code at the time of admission. This type of proxy measure will always have considerable "noise" associated with it, since, for example, a patient in a highly language-isolated area still might not be in a language-isolated household, and vice-versa. The "noise" leads to reduced precision of our impact estimates, making it more difficult to isolate whether the demonstration affected agencies' selection of high per-visit-cost patients.

A second limitation is the possibility of changes taking place between treatment and control agencies over time that would not have been controlled through the DD estimator. Consider, for example, if control agencies experienced a demographic shift in their catchment areas that led to a higher proportion of language-isolated households during the demonstration. Unless treatment agencies experienced a similar change, we might incorrectly conclude that the demonstration had led treatment agencies to avoid patients in language-isolated households. Fortunately, we actually observed very little change over time in our outcome measures, suggesting that such environmental changes were probably not introducing bias into our impact estimates. Nevertheless, we cannot rule out the possibility of this source of bias.

## B. ESTIMATING OVERALL IMPACTS ON PATIENT RETENTION

Unlike the selection analysis, which focused on a single point in time (admission), the retention analysis investigated demonstration impacts over a period of time after admission. To recover unbiased impact estimates, we examined agencies' retention behavior over a fixed time period after the initial home health admission rather than over each episode of care. Below, we discuss the basis for this approach, followed by a summary of the regression models used to recover demonstration effects.

### 1. Choosing the Proper Unit of Analysis

Because treatment agencies might have responded to demonstration incentives by reducing episode lengths, using an episode of care as the unit of analysis could misstate demonstration impacts on patient retention. Consider the following example (which has been contrived for clarity). A typical patient was admitted to a control agency and received home health services for eight months before being transferred to a second agency. If this same patient had been admitted to a treatment agency, the incentives of the demonstration might have led to an earlier discharge (say, at five months), a readmission into a second episode of care, and, finally, a transfer at eight months. Under this scenario, if we used an *episode* as the basic unit of analysis to evaluate the impact of the demonstration, we would conclude that treatment agencies had a lower rate of transfer than control agencies (50 percent, compared to 100 percent). However, if we used the *patient* as the unit of analysis and investigated the rate of transfer within a fixed period from the first admission, we would conclude, more appropriately, that the rate of transfer was unchanged as a result of the demonstration.[7]

In principle, the fixed time period over which we examined agencies' retention behavior could have been any length; however, we faced a trade-off between increasing duration and losing observations because of incomplete data. After taking this into account, we settled on using a "patient-year"--the first 365 days following the home health admission--as the primary unit of analysis. Because we used patients, rather than episodes, as the unit of analysis, we ignored any subsequent episodes that a given patient experienced and defined a patient year *only* from the first admission we observed for the patient. By doing so, our impact estimates were unaffected by the possible correlation between treatment status and subsequent episode starts.

Depending on the date a given agency entered the demonstration, we had a full year of data on service use for anywhere from 12 to 19 months of admissions. The benefit of using all these admissions was that it would improve the precision of the impact estimates. However, because agencies' behavior might have been related to how long they participated in the demonstration, our estimates should be based on a consistent period of participation for all agencies. For our main analysis, we therefore limited our sample to admissions taking place within each agency's first year in the demonstration. In a sensitivity analysis, we then compared impacts during the agency's first and second years in the demonstration in order to see whether or not impacts were stable.

---

[7] If the period we evaluated was less than eight months from admission, we would conclude that the rates of transfer were both zero percent in treatment and control agencies; if the period was eight months or more, we would conclude the transfer rates were both 100 percent. Either way, the treatment-control difference would be zero.

For the majority of "late" agencies, which entered the demonstration on January 1, 1996, we had as little as four months of complete postadmission data for episodes starting in the second demonstration year. We therefore limited the comparison of impacts between the two demonstration years to the first four months (the at-risk period) after an admission. This more limited period should still have allowed us to detect significant differences between the two demonstration years, since, as described in the next chapter, treatment group agencies had the strongest incentive to transfer a patient during this period.

## 2. Regression Models

Since all the outcome variables used to measure patient retention were dichotomous (for example, the likelihood of a transfer by one year postadmission), we used logistic regression models to estimate demonstration impacts. Despite differences in the way the outcome variables were measured, the specification of these logit models was the same as that used to investigate person selection:

$$(4) \quad Probability\ (Y = 1) = \frac{1}{1 + e^{(\alpha + X\beta + \delta T)}},$$

where:

$Y$ = a dichotomous outcome variable, measured from admission through a fixed point in time (for example, Y=1 if the patient experiences a transfer within one year after admission; Y=0 otherwise)

$X$ = a matrix of control variables

$T$ = a binary variable for treatment status that equals one for admissions to treatment agencies and zero for admissions to control agencies

$\alpha$ = the intercept term

$\beta$ = the vector of regression coefficients on the control variables

$\delta$ = a regression coefficient on the variable for treatment status

Recall that, given the nonlinearity of the logit model, the estimated impact of the payment method is not measured directly by the coefficient $\delta$ on the variable for treatment status. Instead, to estimate the demonstration impact on the probability that $Y = 1$, we used the coefficient estimates from the model to generate two predicted probabilities for each observation: one assuming that the observation belonged to the treatment group ($T = 1$), and one assuming that it belonged to the control group ($T = 0$). The impact estimate was given by the average difference between these estimated probabilities. Because the statistical significance of $\delta$ determines whether the odds that $Y = 1$ are significantly different for the treatment and control groups, we used the *p*-value on this parameter to test our hypotheses about differences between the two groups (see Section E).

### 3. Potential Limitations of the Analysis on Patient Retention

Similar to the analysis of cost selection, the greatest limitation in the analysis of patient retention was the lack of available information on our outcomes. Indeed, the claims data not only failed to provide a direct indicator for whether a transfer took place, but they also did not provide the exact date of discharge from admitting demonstration agency. As described in Chapter II, this led us to use a series of outcome measures based on the gap observed between the approximated discharge date (based on the end date of the final claim) and the date of admission to a second agency. Taken together, these measures should provide a reasonably strong test of whether a significant rise in transfers took place in response to the demonstration. However, they cannot provide as precise an impact estimate as a completely accurate measure of patient transfer would have provided.

An additional concern with this analysis, like the others, was the potential for bias resulting from omitted variables in the regression models. Most notably, if preexisting treatment-control differences existed in transfer behavior, it would most likely have biased our impact estimates. Of course, any omitted variable bias should have been mitigated to at least some extent by the combination of random assignment and the use of regression models. Moreover, the results shown in Chapter IV indicated no treatment-control differences in transfer behavior, making it less probable that omitted variables were a major factor in our estimates.[8]

## C. ESTIMATING SUBGROUP IMPACTS

We investigated demonstration impacts among agency subgroups, for two reasons. First, to the extent that impacts existed for the full sample, we would be interested in determining whether they differed between key subgroups of agencies. Second, even if we found no compelling evidence of overall impacts, it is important to determine whether they existed for selected subgroups.

The subgroup analysis used a regression model similar to equation (1). The only difference was that the subgroup model included additional control variables formed by interacting treatment status $T$ with a set of five binary (subgroup) variables defined from agency characteristics. Each subgroup variable included in the regression reflects a key agency characteristic (for example, that the agency was for-profit), and its corresponding (reference) group reflects the absence of this characteristic (for example, that the agency was nonprofit).[9]

---

[8] Given the absence of treatment-control differences, any bias from unobserved factors (such as preexisting transfer behavior) would have had to exactly offset any real impacts arising from the demonstration. While certainly possible, such an outcome is not probable.

[9] As described in Chapter V, the subgroup "pairs" we examined are: (1) for-profit status or nonprofit status, (2) high-use practice patterns versus low-use practice patterns, (3) small size or large size, (4) freestanding or hospital-based, and (5) above or below the base-year cost limits.

The basic regression model for subgroups was given by:

$$(5)\ Y = \alpha + X_1\beta_1 + X_2\beta_2 + \delta T + \sum_{i=1}^{5} \gamma_i\ (X_{1i} * T\ ) + \epsilon,$$

where:

$Y$ = an outcome variable

$X_1$ = a matrix of five subgroup variables, each of which takes on a value of one if the observation belonged to the subgroup, and zero otherwise

$X_2$ = a matrix containing all other control variables

$T$ = a binary variable for treatment status that equals one for admissions to treatment agencies and zero for admissions to control agencies

$\beta_1$ and $\beta_2$ = vectors of regression coefficients on the variables in $X_1$ and $X_2$, respectively

$\gamma_i$ = a regression coefficient on the interaction term between $X_{1i}$ and $T$[10]

$\alpha$ = the intercept term

$\delta$ = a regression coefficient on the variable for treatment status

$\epsilon$ = a random disturbance term assumed to have a mean of zero (conditional on the regressors in the model) that reflects all the unobserved factors affecting the outcome variable

To assess whether the impacts between particular subgroups differed, we examined the statistical significance of the coefficient $\gamma_i$ on the corresponding interaction term. For example, suppose that the *p*-value for the coefficient on the interaction term between an agency's for-profit status and its treatment status is .04 (statistically significant from zero). This test indicates that the effect of prospective payment on that outcome measure differs significantly between for-profit and nonprofit agencies.

---

[10] This interaction term takes on a value of one if the observation belongs to the *i*th subgroup *and* is a treatment admission; and zero otherwise.

We also estimated the effect of prospective payment for each individual subgroup.[11] The impact for a subgroup was estimated by setting the indicator variable for the subgroup appropriately, then obtaining the predicted value of the outcome variable for each observation, first as if it were from a treatment agency ($T = 1$), and then as if it were from a control agency ($T = 0$). The mean difference between these two predicted values provided an unbiased estimate of the impact of prospective payment on the outcome $Y$ for each subgroup. Thus, for the subgroup of agencies with characteristic $j$, the impact estimate was given by:

$$(6) \quad \delta + \sum_{i \neq j} \gamma_i \overline{X_{Ii}} + \gamma_j X_j,$$

where $\overline{X}_{Ii}$ equals the sample mean of the $i$th subgroup variable $X_{Ii}$. The significance of this linear combination was tested to generate the $p$-values for the subgroup impact estimates shown in the first two columns of the Chapter V tables. (See Section E for a further discussion of hypothesis tests.)

## D. ROBUSTNESS TESTS

While our analysis of patient selection and retention used appropriate methods and explored the possible effects of the demonstration in great detail, the findings from the main regression models might have been sensitive to three key factors. First, the impact estimates might have varied over the course of the demonstration. Second, the use of sample weights that equate agencies' representation in the data might have given small agencies undue influence in the analysis. Third, the general existence of outliers, either within or across agencies, might have had undue influence on the analysis. Details of our sensitivity tests are discussed below; in brief, the results from these models (presented in Chapter IV) suggest that our main findings are quite robust.

---

[11] It may not be possible to identify small or moderate impacts for subgroups, as tests of statistical significance lose power as the sample size decreases. The smaller the size of the subgroup (all else equal), the less likely we are to reject the hypothesis that the payment method has no effect, for any given true effect size.

### 1. Comparing Impacts Between Demonstration Years

An important concern is that the impact estimates may have changed over the course of the demonstration. While small changes may not be important, large changes would bring into question the ability to interpret the findings more broadly. We expect that any changes would be due to a delay in agencies' response to the demonstration incentives; however, important market changes (for example, managed care growth) might also cause changes among control agencies that could influence demonstration effects.

To examine whether the impact of the demonstration changed with agencies' length of participation, we combined the available sample from demonstration year 2 with our main analysis sample from demonstration year 1 and estimated a single regression model for each outcome of interest. The specification of this model was a variant on the subgroup model shown in equation (3), where the subgroup of interest was the year of the demonstration rather than some fixed agency characteristics. The model was thus given by:

$$(7)\quad Y = \alpha + X\beta + \delta T + \lambda_1 D + \lambda_2 (T * D) + \epsilon,$$

where:

$Y$ = an outcome variable

$D$ = a dummy variable that equals one if the admission took place in demonstration year 2, and zero if it is from demonstration year 1

$X$ = a matrix of control variables

$T$ = a binary variable for treatment status that equals one for admissions to treatment agencies and zero for admissions to control agencies

$\lambda_1$ = the coefficient on the dummy variable $D$

$\lambda_2$ = the coefficient on the interaction term between the variables $D$ and $T$, where this interaction term equals one for observations reflecting admissions in the second demonstration year to treatment agencies, and zero otherwise

$\delta$ = the intercept term

$\beta$ = the vector of regression coefficients on the control variables

$\delta$ = a regression coefficient on the variable for treatment status

$\epsilon$ = a random disturbance term assumed to have a mean of zero (conditional on the regressors in the model) that reflects all the unobserved factors affecting the outcome variable

As before, the estimated impact of prospective payment in the first demonstration year was given by the coefficient $\delta$. The estimated impact in the second demonstration year was given by the sum of coefficients $\delta$ and $\lambda_2$. The difference between these two impacts, $\lambda_2$, thus measured the change in the effect of prospective payment from the first to the second years of the demonstration.

### 2. Sample Weights

To investigate the sensitivity of our results to the weighting approach, we reestimated the impacts of prospective payment on key outcomes without sample weights. If the impact estimates were similar under the two approaches, it strongly suggests that our results were not overly influenced by a small number of anomalous observations from agencies with few admissions. Thus, the results may be more broadly interpreted for policy purposes. Conversely, although dissimilar results under the two sample weights do not necessarily indicate that the main results were "incorrect," they do suggest the need for additional assessment to determine the most valid inferences for policy.

### 3. Outliers

The distribution of our outcome variables was generally not highly skewed, but it is still important to examine the degree to which our findings on service use were affected by a small number of outlier observations. We used two methods to investigate the effects of outliers. First, to account for outliers at the agency level, we reestimated the selected impacts excluding all admissions from eight agencies--with the two highest and two lowest mean values for a given outcome in both the control group and the treatment group. Because we used sample weights that gave each agency equal weight in these models, the removal of these eight outlier agencies left about 90 percent (71/79) of the weighted sample intact.

In the second approach, we redefined the dependent variable as the natural logarithm of the outcome measure and regressed the log variable on the standard set of explanatory variables shown in equation (1). Although this model effectively reduces the influence given to outliers in the simple linear model, it has two major limitations. First, it might substantially understate the true demonstration effects, if the effects were concentrated among those who would have received the greatest number of visits in the absence of the demonstration. Second, the natural logarithm is undefined for values of zero. As a result, we used this model only to examine the effect of outlier observations on expected utilization.

The log-linear model is given by:

$$lnY = \alpha + X\beta + \delta T + \epsilon, \tag{8}$$

where the variable $lnY$ is the natural log of the outcome variable $Y$, and the remaining variables and parameters are defined as in equation (1). To recover the impact of prospective payment on the outcome $Y$ (rather than on $lnY$) from this model, we must perform the following transformation of the coefficient $\delta$:

$$(9) \quad exp^{\delta} - 1.$$

This transformation yields the impact of prospective payment on the *percentage change* in the outcome $Y$. To recover its impact on the absolute change in $Y$ (equivalent to our impact estimates from prior models), we multiply this value by the sample mean for the outcome $Y$.

## E. STATISTICAL PRECISION AND OTHER MEASUREMENT ISSUES

Previous sections have described the statistical models used to estimate various program impacts. Below, we describe how we measured the statistical significance of these impacts and provide additional detail on the use of statistical weights in the regression models.

### 1. Hypothesis Tests for the Impact Estimates

For each outcome variable that we examined, a two-tailed $t$-statistic tested the null hypothesis that there was no difference between the regression-adjusted population means for treatment and control agencies.[12] The associated $p$-value, which indicates the probability of obtaining a sample estimate of the observed magnitude if the null hypothesis were true, was used to determine whether the demonstration had a measurable impact.

---

[12] Two-tailed tests were used throughout our analysis to avoid confusion and to flag estimates of the "wrong" expected sign that were large enough to be statistically significant. For impacts with the "correct" expected sign, a two-tailed test was less likely than a one-tailed test to reject the hypothesis of no demonstration effect (all else equal).

All *p*-values were based on standard errors adjusted to account for the clustering of patients within agencies and the use of sample weights (see below). Without such an adjustment, the reported *p*-values could have significantly overstated the precision of the reported estimates. This adjustment also accounted for the loss of precision in selected samples that resulted from having multiple observations/admissions for the same patient. A *p*-value of less than .10 was believed sufficient to reject of the null hypothesis and thus provided significant statistical evidence that a demonstration impact exists. At this *p*-value, however, approximately 10 percent of independent tests will show, simply by chance, a statistically significant treatment-control difference when there is no true program effect (known as a Type I error). Therefore, in assessing whether a statistically significant treatment-control difference, especially one with a *p*-value between .05 and .10, should be interpreted as a true program impact, we consider whether the sign and magnitude of the predicted effect are consistent with those for related outcomes.

Throughout the tables of results in Chapters IV and V, we present the means of each outcome variable for the treatment group and the control group alongside the estimated impact, as points of reference. The control group mean is unadjusted and reflected a reasonable estimate of the mean value expected to occur in the absence of the demonstration. The treatment group mean is regression-adjusted by setting it equal to this control group mean minus the (regression-based) impact estimate.

### 2. Design Effects

To draw appropriate inferences about the expected effects of a national prospective payment program, our estimated standard errors must reflect the fact that our observations were clustered in a small number of agencies. The variances of the impact estimates generated from standard statistical packages account for the number of patients included in the sample, treating them as though they were a random sample from an infinite population of patients. However, they do not account for the fact that our sample consisted of patients from a limited number of agencies.

To correct for possible nonindependence of observations within agencies, we used SUDAAN software to obtain the appropriate standard errors for our impact estimates. The SUDAAN calculations also accounted for the greater variance introduced by using sample weights in the regression models.

### 3. Weighting

As noted previously, we weighted observations in the main regression analysis to give each agency equal representation in the analysis. We used this approach for two reasons. First, because the demonstration was implemented at the agency level, the agency was the behavioral unit of interest. Second, the use of weighted data ensures that the impact estimates were not dominated by the experiences of a few large agencies. Preventing this dominance was important because the largest agencies had as much as 100 times the number of observations as the smallest agencies.

For each patient in agency *i*, we constructed the "agency equal" weight ($w_i$) as the ratio of the average number of patient-level observations per agency ($n/k$) to the number in the *i*th agency ($n_i$):

$$(10) \qquad w_i = \frac{n/k}{n_i},$$

where $n$ is the total number of observations for all agencies, and $k$ is the number of agencies. Thus, patient observations for larger-than-average agencies were downweighted, and those for small agencies were given greater weight. As noted above, we also examined selected impact estimates without sample weights as a test of robustness.

# IV. IMPACT OF PROSPECTIVE PAYMENT ON PATIENT SELECTION AND RETENTION

A recent report by Trenholm (2000) finds that per-episode prospective payment markedly lowered the number of visits provided to a home health patient. In this chapter, we examine whether the decline in service use was associated with an overall reduction in access to care, as measured by significant differences in patient selection and retention under the demonstration.

## A. EXPECTED EFFECTS

Under the demonstration, a prospectively paid (treatment) agency had a clear incentive to keep its per-episode costs below the per-episode payment in order to generate agency profits. In response to this incentive, the agency might have chosen to change its patterns of patient selection or retention in order to serve a lower-cost patient mix. Two changes in patient selection might have been particularly advantageous. The first was to reduce the share of admissions expected to be high utilizers of care, while the second was to reduce the share of admissions expected to have relatively high per-visit costs. One strategy agencies could have used to select lower-use patients was to gather detailed information from discharge planners, intake protocols, and other sources in order to sort (and subsequently select) patients based on their expected service use. In addition, the agency might have tried to change its referral base by obtaining new sources with a more favorable (lower-cost) patient mix and perhaps limiting admissions from sources with a more adverse (higher-cost) patient mix. Strategies to avoid selection of patients with high costs per visit could have included reduced admissions from remote areas (where extensive travel time might be necessary) and from non-English-speaking households (where a translator might be needed).

Treatment agencies might also have chosen to change their typical retention practices in response to the demonstration, presumably by transferring high-cost patients as often as possible. Such behavior would have been particularly advisable during the first 120 days after admission, since this was the period during which the agency was "at risk" for all the care that it provided. However, transfer of high-cost or long-stay patients after the first 120 days might also have been advisable if the agency expected its prospective payment for outlier (post-120 day) visits to be below its per-visit costs.

Despite the incentive to select only relatively low-cost patients under the demonstration, several mitigating factors might have caused treatment agencies not to pursue such a strategy. First, it might have been difficult for an agency to collect enough preadmission information to estimate a patient's expected service use. Many agencies, for example, accepted a large proportion of their patients over the phone and would thus have had to adopt major changes in their intake procedures to determine a patient's service needs prior to admission. Second, a strategy to select only relatively low-cost patients might have been a violation of the agencies' mission or the practice norms of the staff. Third, the case-mix adjuster used in the demonstration may have made it difficult for an agency to know whether or not a patient was, in fact, relatively high-cost. For example, although a patient with diabetes might be expected to have higher utilization than a typical patient, agencies' final payment for these patients might be upwardly case-mix-adjusted. This suggests that the most effective method of selection was to obtain healthier patients *within* a given case-mix group--a strategy far less straightforward or easily adopted. Finally, given the strong competition for referrals in most areas, many agencies were probably constrained in adjusting their patient mix in a meaningful way. For example, an agency that tried to admit only healthier patients from a referral source ran the risk of alienating the source and losing it to a competitor. Indeed, even if the referral source provided only high-cost patients, such a loss could be harmful to the agency because it might generate a decline in service volume and thereby raise the agency's per-visit costs.

Similar disincentives existed with respect to transferring patients who proved relatively costly. First, an agency that made a practice of "dumping" its high-cost patients again ran the risk of upsetting referral sources. Second, an agency that transferred a high proportion of its patients might have lost market share, and possibly experienced a decline in its service volume that raised per-visit costs. Third, a strategy to retain only relatively low-cost patients might have violated an agency's mission or the norms of its staff.

Previous findings provide ambiguous evidence about whether the demonstration altered agencies' selection or retention practices. Trenholm (2000) found that treatment agencies sharply reduced both episode lengths and visit frequency in response to the demonstration, resulting in a substantial decline in overall utilization per patient. One interpretation of this finding is that agencies chose to curtail service provision in order to reduce their per-episode costs, rather than alter their practices of selection or retention. Indeed, while information gathered during site visits strongly suggested that treatment agencies had cut back their visits per patient, it provided little evidence that agencies changed their referral sources, intake methods, or transfer patterns to obtain a more favorable patient mix. Nevertheless, an alternative interpretation of these findings is that they show just how strongly treatment agencies were driven to cut back per-episode costs in response to the demonstration, suggesting that they would make important changes in patient selection or retention as well. Moreover, the observed reductions in episode lengths were entirely consistent with possible efforts to transfer high-cost patients frequently to obtain a more favorable patient mix.

In summary, given the contradictory incentives faced by demonstration agencies, it is not possible to predict how the demonstration affected patient selection and retention without an empirical analysis that isolates agencies' behavior in this regard. In the rest of this chapter, we present findings from such an analysis--first, with respect to patient selection, and, second, with respect to patient retention. In Chapter V, we present results from a subgroup analysis that explores whether specific agency types responded differently to the demonstration incentives.

## B. IMPACTS ON PATIENT SELECTION

Based on an analysis of patient characteristics highly predictive of service use during the at-risk period, we find no evidence to suggest that the demonstration affected agencies' selection of patients (Table IV.1). For the 12 characteristics evaluated, only 1--"treatment can likely be taken over"--displayed a significant impact consistent with the incentive to select lower-use patients. Given the lack of impacts elsewhere, this one significant outcome does not provide compelling evidence that the demonstration affected patient selection; in fact, there is reason to believe that it is proxying for the effects of the demonstration elsewhere. Namely, because treatment agencies probably were focusing more on finding alternative care sources in order to cut their service provision, this isolated impact is most likely picking up agencies' plans to lower service use rather than measuring a change in selection practices. Indeed, this interpretation is consistent with both the sharp reductions in service lengths found under the demonstration (Trenholm 2000) and information from site visits suggesting that treatment agencies tried to get patients and their family members more involved in care provision.

## TABLE IV.1

## PERCENTAGE OF PATIENTS ADMITTED WITH SELECTED CHARACTERISTICS THAT PREDICT A SIGNIFICANT CHANGE IN SERVICE USE DURING THE AT-RISK PERIOD

| Predictors of Service Use[a] | Treatment Group Mean[b] (Percentage) | Control Group Mean (Percentage) | Difference[b] (*p*-value)[c] |
|---|---|---|---|
| Has Urinary Incontinence | 8.1 | 8.5 | -0.4<br>(.43) |
| Has Decubitus, Stage 3 or 4 | 3.6 | 3.6 | -0.0<br>(.98) |
| Had Cebrovascular Accident (CVA) | 15.3 | 14.8 | 0.5<br>(.64) |
| Has Diabetes | 21.0 | 20.2 | 0.8<br>(.52) |
| Requires Complex Wound Care | 6.3 | 7.3 | -1.0<br>(.22) |
| Surgical Wound Not Healing | 1.2 | 1.2 | -0.0<br>(.78) |
| Unstable at Admission | 51.5 | 55.5 | -4.0<br>(.21) |
| Needs Assistance in Ambulation | 20.9 | 19.5 | 1.4<br>(.25) |
| Unable to Do Light Housekeeping | 67.8 | 68.0 | -0.2<br>(.92) |
| Life Expectancy Is More Than Six Months | 92.5 | 93.4 | -0.9<br>(.41) |
| Paid Help Is Available | 16.6 | 15.9 | 0.7<br>(.65) |
| Treatment Can Likely Be Taken Over (by Patient, Friends, Neighbor, or Paid Help) | 72.5 | 66.7 | 5.8***<br>(.01) |
| **Sample Size** | | | |
| **Observations** | **32,118** | **22,235** | |
| **Agencies** | **45** | **39** | |

SOURCE: Predicted utilization based on patient-level information from Medicare claims, QA admission forms, UB-92 remarks fields (on the admission claims), Medicare enrollment files, and agency cost reports.

NOTE: Each observation in the sample reflects a patient's episode start in a home health agency participating in the demonstration. The observations have been weighted to give agencies equal representation in the calculations.

[a] These variables predicted a change in utilization during the at-risk period of at least 10 percent relative to the patient average. The final two measures, "paid help is available" and "treatment can likely be taken over," predicted lower use; the remaining measures all predicted higher use.

[b] Regression-adjusted (through ordinary logit models) to control for preexisting differences between treatment and control agencies.

[c] The *p*-value corresponds to a test of whether the treatment-control difference (impact) is statistically different from zero. It is based on standard errors inflated to account for the effects of clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

Two composite measures of patient's expected service use support the findings of no significant impacts on agencies' overall selection practices (Table IV.2, full sample results).[1] Under the primary measure, the estimated difference in expected service use among treatment and control group patients was small, insignificant, and with the wrong sign to indicate favorable selection by treatment agencies. Similarly, impact estimates with the secondary measure, which included indicators of planned treatments to predict service use, were small and insignificant. The small size of the latter impact actually is somewhat surprising, since we viewed it as an upper bound on how strongly agencies would favorably select in response to the demonstration.

When we separated the sample by first and later episodes (Table IV.2, split sample results), we did find significant differences in the expected use of patients readmitted into a later episode of care. Under the primary measure, this difference was modest ($179, or about five percent) but nevertheless significant, whereas it was considerably larger under the secondary measure. One explanation for this result, consistent with the incentives of the demonstration, is that treatment agencies reduced their readmission of high-need patients in an effort to obtain a more favorable patient mix. Indeed, such an approach might be particularly easy to pursue for this subsample, since an agency would have very detailed knowledge of patients admitted previously. An alternative explanation, however, is that both our measures of expected service use are biased for later episodes because some of the patient characteristics on which they are based are correlated with an agency's treatment status. For

---

[1] Our primary measure predicts a patient's home health use during the at-risk period, based on a large set of observed characteristics at admission. A secondary measure uses these same characteristics to predict utilization, as well as selected treatments planned by the admitting agency during the episode of care. By including these planned treatments, we improve the relative accuracy of this measure; however, we might also bias any treatment-control differences (in favor of showing an impact) if treatment agencies reduced the treatments planned for a given patient in response to the demonstration. See Appendix C for further details on how these measures were constructed. Limitations with these measures are discussed in Chapter III.

### TABLE IV.2

### PREDICTED LEVEL OF HOME HEALTH UTILIZATION DURING THE AT-RISK PERIOD, BASED ON PATIENT CHARACTERISTICS OBSERVED AT ADMISSION

| Measure of Predicted Utilization[a] | Treatment Group Mean[c] (Dollars) | Control Group Mean (Dollars) | Difference[c] (*p*-value)[d] |
|---|---|---|---|
| **All Episodes** | | | |
| Primary Measure | 3,289 | 3,220 | 69<br>(.33) |
| Secondary Measure | 3,185 | 3,217 | -32<br>(.67) |
| **First Episodes Only**[b] | | | |
| Primary Measure | 3,273 | 3,184 | 89<br>(.20) |
| Secondary Measure | 3,189 | 3,191 | -2<br>(.98) |
| **Later Episodes Only**[b] | | | |
| Primary Measure | 3,304 | 3,483 | -179**<br>(.03) |
| Secondary Measure | 3,088 | 3,408 | -320***<br>(.01) |
| **Sample Sizes** | | | |
| **Observations** | **32,118** | **22,235** | -- |
| **Agencies** | **45** | **39** | -- |

SOURCE: Predicted utilization (during the first 120 days after admission) based on patient-level information from Medicare claims, QA admission forms, UB-92 remarks field (on the admission claim), Medicare enrollment files, and agency cost reports.

NOTE: Each observation in the sample reflects a patient's episode start in a home health agency participating in the demonstration. The observations have been weighted to give agencies equal representation in the calculations.

[a] The primary measure is based on all available information on patients at admission that might be predictive of their utilization *except* information on planned treatments (available from the QA admission forms). The secondary measure includes the planned treatment information as well. While highly predictive of service use, the treatments planned for a patient might be affected by an agency's treatment status, possibly leading to upward bias in the estimated demonstration impact. The impact estimate under the secondary measure should thus be viewed as an upper bound.

[b] The subsample of the first episodes includes 28,415 observations in 45 treatment agencies and 19,682 observations in 39 control agencies--roughly 88 percent of all observations in the full sample. The subsample of later episodes includes 3,661 observations in 36 treatment agencies and 2,533 observations in 33 control agencies. Agencies with fewer than 10 later episodes were dropped from this subsample; thus, the two subsamples add to slightly less than the total.

[c] Regression-adjusted (through ordinary least squares model) to control for preexisting difficulties between treatment and control agencies.

[d] The *p*-value corresponds to a test of whether the treatment-control difference (impact) is statistically different from zero. It is based on standard errors inflated to account for the effects of clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

example, the number of home health visits prior to admission is a significant, positive predictor of service use and thus was included in our measures. But, *among readmissions to treatment agencies*, prior visits might be significantly lower because of cutbacks made in response to the demonstration, thereby leading to downward bias in our measures of expected service use for this group.

Additional estimates (not shown) prove more consistent with estimation bias as the explanation for the significant results in Table IV.2. In particular, when we used a measure of expected service use that excluded the most likely endogenous variables, the treatment-control difference among later episodes fell appreciably to only three percent (-$110) and became statistically insignificant (*p*-value of 0.14).[2] In addition, when we reexamined the outcomes in Table IV.1 for later episodes only, our results were similar to those shown in the table for the full sample and not at all suggestive of favorable selection by treatment agencies. Therefore, while we cannot ignore the possibility that agencies might have avoided higher-cost readmissions in response to the demonstration, the evidence to support this hypothesis is weak.

A more refined approach that treatment agencies might have used to improve their case mix was to select low-cost patients from *within* the individual Home Health Utilization Groups (HHUGs). As described in Appendix A, these HHUGs were used to retrospectively adjust agencies' episode payment to reflect changes in their case mix between the predemonstration and demonstration periods. By selecting lower-cost patients from within a given HHUG, agencies would have maximized their potential gain from altering their case mix, while at the same time such behavior might have been overlooked in our analysis of overall patient mix reported in Tables IV.1 and IV.2.[3]

---

[2] The variables excluded were: whether treatment can likely be taken over; number of home health visits in the six months before admission; total Medicare reimbursement in the six months before admission; whether admitted from hospital; length of pre-admission hospital stay; and whether patient had a SNF stay within 14 days before admission.

Despite the potentially strong incentive for this type of selection, however, we find no evidence that treatment agencies altered their patient mix within HHUGs in response to the demonstration (Table IV.3). Within each HHUG evaluated, the treatment-control differences in expected use were small (the largest was less than three percent) and insignificant, and a number of these had the wrong sign to suggest that the demonstration led to favorable selection. In addition, estimates of the proportion of patients in each of the various HHUGs (not shown) displayed no significant changes as a result of the demonstration. This finding is perhaps not surprising, since agencies could not easily have anticipated the financial implications of altering their case-mix distribution; nevertheless, it reinforces how little evidence exists of changes in patient selection under the demonstration.

An entirely different approach to selection that treatment agencies might have adopted was to avoid patients whose costs per visit were higher than normal because of their location, language isolation, or other factors. Finding evidence of this type of selection is more difficult because the available measures are less precise than those used to examine expected utilization. Nevertheless, based on our best available measures, we find no indication that treatment agencies avoided patients who might have had higher per-visit costs. For instance, based on difference-in-difference models shown in Table IV.4, we find no effect from the demonstration on the percentage of patients admitted from highly rural areas or from areas with a high degree of language isolation. Similar

---

[3] As an example of how this "within-HHUG" selection could be missed in an analysis of overall selection behavior, consider an agency that admitted fewer patients belonging to the lowest -cost HHUG by identifying and avoiding those expected to be high utilizers of care *within this HHUG.* This within-HHUG selection could have led to a substantial windfall for the agency, even if it led to no change (or even an increase) in the expected utilization of its average patient.

## TABLE IV.3

## PREDICTED LEVEL OF HOME HEALTH UTILIZATION WITHIN THE MAJOR CASE-MIX CATEGORIES

| Home Health Utilization Group (HHUG)[a] | Treatment Group Mean[b] (Dollars) | Control Group Mean (Dollars) | Difference[b] (*p*-value)[c] |
|---|---|---|---|
| HHUG 3 | 2,449 | 2,463 | -14<br>(.86) |
| HHUG 7 | 4,773 | 4,854 | -81<br>(.45) |
| HHUG 9 | 3,967 | 4,037 | -70<br>(.20) |
| HHUG 12 | 2,935 | 2,951 | -16<br>(.83) |
| HHUG 13 | 2,037 | 2,047 | -10<br>(.85) |
| HHUG 15 | 4,030 | 3,925 | 105<br>(.18) |
| HHUG 16 | 4,484 | 4,453 | 31<br>(.64) |
| HHUG 17 | 4,354 | 4,305 | 49<br>(.43) |
| HHUG 18 | 3,296 | 3,234 | 62<br>(.33) |
| **Sample Size (per HHUG evaluated)** | | | |
| **Mean Observations** | **3,019** | **2,112** | -- |
| **Minimum Observations** | **545** | **424** | -- |
| **Maximum Observations** | **9,540** | **6,478** | |

SOURCE: Predicted utilization (during the first 120 days after admission) based on patient-level information from Medicare claims, QA admission forms, UB-92 remarks fields (on the admission claim), Medicare enrollment files, and agency cost reports.

NOTE: Each observation in the sample reflects a patient's episode start in a home health agency participating in the demonstration. The observations have been weighted to give agencies equal representation in the calculations.

[a] Only HHUGs that included more than two percent of the sample and more than half the agencies were investigated. Predicted utilization level is based on the primary measure (that excludes information on planned treatments) and reflects service during the first 120 days after admission (at-risk period).

[b] Regression-adjusted (through ordinary least squares models) to control for preexisting differences between treatment and control agencies.

[c] The *p*-value corresponds to a test of whether the treatment-control difference (impact) is statistically different from zero. It is based on standard errors inflated to account for the effects of clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

## TABLE IV.4

## CHANGE IN THE PROPORTION OF PATIENTS SERVED WITH POTENTIALLY HIGH PER-VISIT COSTS (FROM THE PREDEMONSTRATION TO THE DEMONSTRATION)

| | Treatment Group Mean (Percentage) | Control Group Mean (Percentage) | Difference ($p$-value)[c] |
|---|---|---|---|
| **Proportion of Patients Served in Highly Rural Locations**[a] | | | |
| Demonstration | 7.0 | 8.8 | -1.8 |
| Predemonstration | 7.1 | 8.4 | -1.3 |
| Difference | -0.1 | 0.4 | -0.5 [Impact]<br>(.54)[c] |
| **Proportion of Patients Served in Highly Language-Isolated Locations**[b] | | | |
| Demonstration | 3.1 | 0.6 | 2.5 |
| Predemonstration | 3.4 | 0.5 | 2.9 |
| Difference | -0.3 | 0.1 | -0.4 [Impact]<br>(.20)[c] |
| **Sample Size (Agencies)** | **46** | **41** | |

SOURCE: Medicare enrollment data (for patient location) and 1990 U.S. Census Bureau (for location characteristics).

[a] Patients residing in a zip code that is 100 percent rural and outside a Metropolitan Statistical Area (MSA).

[b] Patients residing in a zip code where at least 30 percent of the households are non-English-speaking.

[c] The $p$-value corresponds to a test of whether the treatment-control difference in the predemonstration to demonstration change (impact) is statistically different from zero.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

results were found under several alternative measures of high per-visit costs (previously listed in Table II.5), further suggesting that the demonstration had no significant effect on patient selection.

## C. IMPACTS ON PATIENT RETENTION

Earlier findings from Trenholm (2000) showed a dramatic reduction in service lengths attributable to the demonstration.[4] Although this reduction could have been due partly to higher transfer rates among treatment agencies, we find no evidence to support this conclusion. As Table IV.5 shows, the transfer rate at various time periods after admission was small in both treatment and control group agencies, and none of the differences between the two groups were statistically significant.[5] The lack of difference within the first few months after admission is perhaps surprising, since treatment agencies discharged a significantly higher share of their patients by this point. For example, by two months after admission, 57 percent of patients in treatment agencies had been discharged from their episode, compared to 49 percent of those in control agencies (Trenholm 2000). Yet, the percentage of patients transferred by two months after admission was identical between treatment and control group agencies, at only 1.8 percent.

---

[4] Episodes of short, medium, and long duration all declined significantly, with the average episode length falling by roughly 25 percent. In addition, the proportion of patients receiving a visit from the admitting agency after the (initial) at-risk period fell by more than 20 percent.

[5] Because we could not observe a patient transfer directly from the claims data, we had to impute whether a transfer had taken place by assessing the gap between the admitting agency discharge and an other-agency admission. Under our "base" transfer measure (reported in Table IV.5), we assumed that a transfer had taken place if there was less than four days between discharge and an other-agency admission, or if there were less than eight days between an other-agency claim starting before the discharge date and a second claim starting after the discharge date. We also examined a "less restrictive" measure (see Table IV.6) that allows these respective gaps be less than 8 days and less than 15 days, and a "least restrictive" measure that lets these gaps be less than 15 days and less than 32 days.

## TABLE IV.5

## CUMULATIVE RATE OF TRANSFER DURING THE YEAR AFTER ADMISSION

| Percent of Patients Transferred Between Their First Admission (Episode Start) and: [a] | Treatment Group Mean[b] (Percentage) | Control Group Mean (Percentage) | Difference[b] ($p$-value)[c] |
|---|---|---|---|
| **At-Risk Period** | | | |
| 30 Days | 1.4 | 1.2 | 0.2<br>(.28) |
| 60 Days | 1.8 | 1.8 | -0.0<br>(.98) |
| 90 Days | 1.9 | 2.1 | -0.2<br>(.53) |
| 120 Days | 2.2 | 2.3 | -0.1<br>(.71) |
| **Post-At-Risk Period** | | | |
| 150 Days | 2.3 | 2.4 | -0.1<br>(.82) |
| 180 Days | 2.6 | 2.5 | 0.1<br>(.82) |
| 240 Days | 2.7 | 2.7 | -0.0<br>(.96) |
| 300 Days | 2.8 | 2.9 | -0.1<br>(.69) |
| 1 Year | 2.9 | 3.2 | -0.3<br>(.41) |
| **Sample Size** | | | |
| **Observations** | **26,384** | **25,749** | -- |
| **Agencies** | **39** | **40** | -- |

SOURCE: Medicare claims data.

NOTE: Each observation in the sample reflects a patient's initial admission to a home health agency during its first year of participation in the demonstration. The observations have been weighted to give agencies equal representation in the calculations.

[a] A transfer is assumed if there are 0 to 3 days between discharge and other-agency admission, or 0 to 7 days between an other-agency claim starting before the discharge date and a second claim starting after the discharge date.

[b] Regression-adjusted (through logit regression models) to control for preexisting differences between treatment and control agencies.

[c] The *p*-value corresponds to a test of whether the treatment-control difference (impact) is statistically different from zero. It is based on standard errors inflated to account for the effects of clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

TABLE IV.6

RATE OF TRANSFER AND OTHER AGENCY INVOLVEMENT BY ONE YEAR AFTER ADMISSION, UNDER ALTERNATIVE MEASURES

| | Treatment Group Mean[a] (Percentage) | Control Group Mean (Percentage) | Difference[a] (*p*-value)[b] |
|---|---|---|---|
| **Transfer Measures**[c] | | | |
| Base Measure (transfer very likely) | 2.9 | 3.2 | -0.3<br>(.41) |
| Less restrictive measure | 4.2 | 4.2 | -0.0<br>(.91) |
| Least restrictive measure | 5.2 | 5.2 | 0.0<br>(.99) |
| **Other-Agency Involvement Measures** | | | |
| Any other-agency visits prior to discharge | 1.6 | 2.0 | -0.4<br>(.14) |
| Any visits predischarge and/or transfer (least restrictive measure) | 6.0 | 6.3 | -0.3<br>(.55) |
| **Sample Size** | | | |
| **Observations** | **26,384** | **25,749** | -- |
| **Agencies** | **39** | **40** | -- |

SOURCE: Medicare claims data.

NOTE: Each observation in the sample reflects a patient's initial admission to a home health agency during its first year of participation in the demonstration. The observations have been weighted to give agencies equal representation in the calculations.

[a] Regression-adjusted (through logit regression models) to control for preexisting differences between treatment and control agencies.

[b] The *p*-value corresponds to a test of whether the treatment-control difference (impact) is statistically different from zero. It is based on standard errors inflated to account for the effects of clustering and weighting.

[c] Base measure: 0 to 3 days between discharge and other-agency admission, or 0 to 7 days between an other-agency claim that starts before the discharge date and a second claim after the discharge date. Less restrictive measure: 0 to 7 days between discharge and other-agency admission, or 0 to 14 days between an other-agency claim that starts before the discharge date and a second claim after the discharge date. Least restrictive measure: 0 to 14 days between discharge and other-agency admission, or 0 to 31 days between an other-agency claim that starts before the discharge date and a second claim after the discharge date.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

Estimates under alternative measures of transfer and other-agency "involvement" further indicate that the demonstration had no adverse effects on patient retention (Table IV.6).[6] For example, based on two less restrictive measures of an agency transfer, the transfer rates between treatment and control group agencies remained nearly identical. In addition, measures that accounted for other involvement by nonadmitting agencies in the episode of care showed only small differences between the treatment and control groups, and these differences were not in a direction consistent with the incentives to obtain a more favorable patient mix.

## D. ROBUSTNESS

Several robustness tests were performed to ensure that our estimates of demonstration impacts were not heavily dependent on particular observations or statistical procedures. First, we investigated impacts on patient selection and retention using regression models without sample weights. Such unweighted estimates place relatively more emphasis on the behavior of large agencies rather than placing equal emphasis on all agencies. Second, we examined the sensitivity of our results to outliers: extreme values for particular patients and agencies. Third, we investigated whether impacts differed between the first and second years of the demonstration.

Each robustness test focused on six key measures of patient selection and retention. The patient selection measures we examined were: expected utilization (primary measure), proportion of patients with diabetes, proportion with decubiti, and proportion with paid help available. We focused on the three latter conditions (as opposed to any others listed in Table IV.1) because they were especially strong predictors of service use, they occurred with some frequency, and they were not difficult to identify prior to admission. The retention measures we focused on were the rate of transfer under the base measure and the rate under the least restrictive measure. For the robustness tests associated with sample weights and outliers, we examined the transfer rate at one year postadmission, a point at which most patients had experienced at least one discharge from the admitting demonstration agency. For the comparison of demonstration years, however, we examined these rates at only four months postadmission (the end of the at-risk period), due to data limitations.[7] This shorter period should still be adequate to test for changes over time in retention practices, since the strongest incentives to transfer under the demonstration were within the at-risk period.

---

[6] The variable for "other-agency involvement" accounts for any episode with a claim from a second agency with a start date before the date of discharge. About half these cases were also classified as transfers (under the base measure) as an additional claim appeared soon after admission. For at least one-quarter of the cases, however, there clearly was no transfer, as all claims appeared well before discharge (see Chapter II for additional details).

### 1. Findings

As Table IV.7 shows, the impact estimates without sample weights or with controls for outlier effects (see Columns 1 and 2) are entirely consistent with the main findings. For each outcome in the table, the estimated differences between treatment and control agencies typically were small and statistically insignificant. The most notable differences were in the proportion of patients admitted with diabetes and the rate of transfer under the basic measure; however, these differences actually were in the wrong direction to suggest that treatment agencies had gamed the payment system to obtain a more favorable patient mix.

The comparison of impacts over time reveals little evidence that treatment agencies altered their selection or retention patterns in either the first or second year of the demonstration (Table IV.8). We do find a positive and significant difference in the proportion of patients with paid help available

---

[7] As described in Chapter II, we observed agency behavior for most year 2 admissions for less than 12 months, due to the cutoff date for our data collection. However, we observed at least four months of agency behavior for all these admissions, including whether a transfer might have taken place.

## TABLE IV.7

## COMPARISON OF IMPACTS ON KEY PATIENT SELECTION AND RETENTION MEASURES UNDER ALTERNATIVE EMPIRICAL SPECIFICATIONS

| | Impacts Under Alternative Models[c] | | | Main Analysis[c] |
|---|---|---|---|---|
| Outcome | Impact Without Sample Weights ($p$-value)[d] | Impact with 80 Percent Agency Subsample ($p$-value)[d] | Impact with Log-Linear Specification ($p$-value)[d] | Impact ($p$-value)[d] |
| **Selection Measures[a]** | | | | |
| Predicted Utilization, Primary Measure (Dollars) | 12 | 95 | 62 | 69 |
| | (.89) | (.14) | (.36) | (.33) |
| Had Diabetes Condition (Percentage) | 1.7* | 1.6 | [e] | 0.8 |
| | (0.9) | (.10) | | (.52) |
| Has Serious Decubiti (Percentage) | 0.1 | 0.7* | [e] | -0.0 |
| | (.92) | (.10) | | (.98) |
| Has Paid Help Available (Percentage) | 0.7 | 1.5 | [e] | 0.7 |
| | (.65) | (.25) | | (.65) |
| **Retention Measures[b]** | | | | |
| Transfer by One Year Postadmission, Base Measure (Percentage) | -0.6** | -0.6** | [e] | -0.3 |
| | (.02) | (.05) | | (.41) |
| Transfer by One Year Postadmission, Less Restrictive Measure (Percentage) | -0.3 | -04 | [e] | -0.0 |
| | (.28) | (.30) | | (.91) |
| Sample Size | | | | |
| **Observations (selection)** | **54,357** | **Variable[f]** | **54,357** | **54,357** |
| **Observations (retention)** | **52,133** | **Variable[f]** | **52,133** | **52,133** |

SOURCE: Selection measures obtained from Medicare claims, QA admission forms, UB-92 remarks fields, Medicare enrollment files, and agency cost reports. Retention measures obtained from Medicare claims data.

NOTE: Each observation in the sample reflects a patient's admission to (episode start in) a home health agency participating in the demonstration. The selection sample includes all available episodes. The retention sample includes only initial episodes in the first demonstration year.

[a] Measured at home health admission. The predicted utilization is based on the primary measure and reflects service use during the first 120 days after admission (at-risk period). Serious decubiti defined as stage 3 or stage 4.

[b] Base transfer measure: 0 to 3 days between discharge and other-agency admission, or 0 to 7 days between an other-agency claim that starts before the discharge date and a second claim after the discharge date. Less restrictive measure: 0 to 7 days between discharge and other-agency admission, or 0 to 14 days between an other-agency claim that starts before the discharge date and a second claim after the discharge date.

[c] Impacts regression-adjusted to control for preexisting differences between treatment and control agencies.

[d] The $p$-value corresponds to a test of whether the impact is statistically different from zero. It is based on standard errors inflated to account for the effects of clustering and any sample weighting.

[e] Outcome is not estimated by log-linear model because outcome variable is binary.

[f] Total sample sizes varied slightly, depending on the agencies dropped for a given outcome. The selection samples were roughly 51,300; the retention samples were roughly 50,700.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

TABLE IV.8

COMPARISON OF IMPACTS ON KEY PATIENT SELECTION AND RETENTION MEASURES BETWEEN YEAR 1 AND YEAR 2 OF THE DEMONSTRATION

| | Impact in Demonstration Year 1[c] (*p*-value)[d] | Impact in Demonstration Year 2[c] (*p*-value)[d] | Difference in Impacts, Year 2-Year 1[c] (*p*-value)[d] |
|---|---|---|---|
| **Selection Measures[a]** | | | |
| Predicted Utilization (Dollars) | 32<br>(.66) | 106<br>(.20) | 73<br>(.32) |
| Had Diabetes Condition (Percentage) | 0.8<br>(.31) | -0.5<br>(.66) | -1.3<br>(.26) |
| Has Serious Decubiti (Percentage) | 1.1<br>(.53) | 0.5<br>(.70) | -0.6<br>(.71) |
| Has Paid Help Available (Percentage) | -1.7<br>(.39) | 2.4<br>(.15) | 4.1**<br>(.03) |
| **Retention Measures[b]** | | | |
| Transfer by End of At-Risk Period, Base Measure (Percentage) | -0.1<br>(.80) | 0.3<br>(.45) | 0.4<br>(.33) |
| Transfer by End of At-Risk Period, Less Restrictive Measure (Percentage) | 0.2<br>(.46) | 0.7<br>(.14) | 0.5<br>(.35) |
| Sample Size | | | |
| **Observations (selection)** | **21,241** | **31,883** | -- |
| **Observations (retention)** | **52,133** | **32,918** | -- |

SOURCE: Selection measures obtained from Medicare claims, QA admission forms, UB-92 remarks fields, Medicare enrollment files, and agency cost reports. Retention measures obtained from Medicare claims data.

NOTE: Each observation in the sample reflects a patient's admission to (episode start in) a home health agency during a given demonstration year. The selection sample includes all available episodes. The retention sample includes first episodes only.

[a] Measured at home health admission. The predicted utilization is based on the primary measure and reflects service use during the first 120 days after admission (at-risk period). Serious decubiti defined as stage 3 or stage 4.

[b] Retention measures investigate only the first 120 days (at-risk) period due to data limitation in year 2. Base transfer measure: 0 to 3 days between discharge and other-agency admission, or 0 to 7 days between an other-agency claim that starts before the discharge date and a second claim after the discharge date. Less restrictive measure: 0 to 7 days between discharge and other-agency admission, or 0 to 14 days between an other-agency claim that starts before the discharge date and a second claim after the discharge date.

[c] Regression-adjusted to control for preexisting differences between treatment and control agencies.

[d] The *p*-value corresponds to a test of whether the impact is statistically different from zero. It is based on standard errors inflated to account for the effects of clustering and sample weights.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

between the two demonstration years; however, this difference was created as much by a negative impact in year 1 as a positive impact in year 2. Indeed, when evaluating the year 2 impact alone, we find a more modest and insignificant increase of 2.4 percentage points. This difference thus provides little evidence of an impact on selection in year 2, particularly in the context of the remaining impacts, which were small, insignificant, and inconsistently signed.[8]

In addition to the robustness tests summarized in Tables IV.7 and IV.8, we would have liked to test formally whether our impact estimates were biased by the loss of observations from the various samples. Unfortunately, such tests were not possible because of the limitations of the data; the evidence we do have, however, suggests that our findings were not at all influenced by missing observations. Easily the most significant sample loss was the removal of nearly half the observations from the use selection sample due to missing quality assurance (QA) information. Since most of these missing data were due to the QA admission form not being collected early in the demonstration--a random "event" that affected treatment and control agencies similarly--we have no reason to believe that it affected our main findings. Moreover, since we might have expected the demonstration to have its weakest effect in the earliest months of the demonstration, any bias from losing these early observations should have been in the direction of overstating the demonstration's effect.

The other important loss of sample was the exclusion of eight agencies from the retention sample because of missing bill information.[9] In results not shown, we find that the impact estimates when including these agencies in the sample were unchanged from those presented earlier. Although this similarity might have been due to undercounts in transfer rates for treatment agencies with missing bill information, it nevertheless provides no evidence that the exclusion of these agencies biased our main results. Moreover, in light of the findings of "no impacts" from the selection analysis (which included these agencies), there is little reason to believe that the retention impacts would have been different if we had complete billing data on all agencies.

---

[8] Impact estimates based on the other major predictors of service use (not shown) also showed no evidence of changes in patient selection between the first and second years of the demonstration.

[9] See Trenholm (2000) for a detailed discussion of this issue.

### 2. Overview

In conclusion, the robustness tests confirm the stability of our main finding that the demonstration did not significantly affect agencies' practices of selection or retention. None of our results were affected by the use of sample weights, by outlier values among agencies or individual patients, or by the timing of a patient's admission during the demonstration. In addition, while we have less definitive evidence to show that the loss of observations from our analysis samples did not affect our impact estimates, we have absolutely no evidence to suggest that it did.

# V. SUBGROUP IMPACTS

Agencies may have responded differently to the demonstration incentives, depending on their individual characteristics. To investigate this possibility, we reestimated key overall demonstration impacts for a series of agency-level subgroups most likely associated with differences in agency behavior.

Given the absence of impacts for the full sample of agencies (see Chapter IV), the key issue for this subgroup analysis is whether selection or retention effects were evident for any specific agency types. If we find no evidence of impacts across most or all agency subgroups, it strengthens the validity of the main findings, suggesting that they may be more easily applied to agencies outside the demonstration. A related issue is whether demonstration impacts differed between specific subgroups of agencies (for example, between nonprofit and for-profit agencies). To the extent that differences existed, it suggests that prospective payment offered agencies varying opportunities (or incentives) to alter patient selection or retention, an important consideration when implementing a nationwide payment system.

The five agency-level characteristics we examined were their for-profit status, affiliation, size, predemonstration practice patterns, and chain membership. From these respective characteristics, we defined the following subgroup pairs:

1. Proprietary and nonprofit
2. Hospital-based and freestanding
3. Small and large--small agencies provided fewer than 30,000 visits in the base year, large agencies provided more than 30,000 visits
4. "High-use" practice pattern and "low-use" practice pattern. These subgroups are defined from a case-mix-adjusted index of agencies' service provision in the base year (see

Appendix B for details). High-use agencies have an index value above the median value; low-use agencies are at or below the median value.

5. Chain member and not a chain member

For each subgroup, we estimated program impacts on key selected measures of selection and retention. We then conducted statistical tests to determine whether (1) the program impact for a given subgroup was significantly different from zero, and (2) a significant difference existed in the impacts between subgroup pairs. The impact estimates and corresponding statistical test results are displayed in tables in this chapter. Additional information, including the underlying treatment and control group means, is provided in a set of companion tables in Appendix D.

## A. EXPECTED IMPACTS

We believe that proprietary agencies were more likely than nonprofit agencies to change their patient selection and retention in response to the demonstration. Like for-profit agencies, nonprofits had an incentive to reduce per-episode costs and generate a surplus that could be applied toward charity work, capital improvements, or other objectives. However, nonprofits may have resisted this incentive because their mission was to provide service to all patients eligible for home health care. In fact, in light of the loss protection under the demonstration, nonprofit agencies could have used prospective payment as an opportunity to expand care to higher-cost groups rather than limit it.

Because a large fraction of the admissions at hospital-based agencies typically come from their hospital affiliate(s), these agencies probably had a much greater opportunity to identify and ultimately select a lower-cost mix of patients. In fact, during our site visits, some freestanding agencies complained that hospitals were in a far stronger position under the demonstration because they could easily "skim" patients expected to require the least care.

Smaller agencies (those under 30,000 visits per year) may have responded differently to the demonstration incentives than large agencies, but the direction of this difference is uncertain. On the one hand, smaller agencies may have found it easier to communicate with their staff and implement large-scale changes in their selection or retention practices. On the other hand, the smaller agencies had a more limited referral base, making them potentially more reluctant (or less able) to make substantial changes in their patient mix.

Agencies with relatively low-use practice patterns might have been more likely to change their selection or retention because they were less able to cut episode costs in other ways. These agencies more effectively controlled their service provision prior to the demonstration and thus may have been less able to make large-scale reductions in visits without adversely affecting patient outcomes. When faced with the incentives of prospective payment, low-use agencies might therefore have more aggressively pursued strategies to obtain a lower-cost patient mix.

Under the demonstration, agencies that were chain members had the unique ability to transfer patients without losing them to a competitor. This, in turn, might allow for the creation of a lower-cost mix without the potential for a decline in market share or the loss of referrals. We therefore expect that chain agencies were among the most likely to raise their transfer rates in an effort to obtain a more favorable patient mix.[1]

---

[1] This advantage would not be afforded to chain agencies under a national payment system, since all member agencies would be subject to the same payment rules. Chain agencies' response to the demonstration thus does not serve as a very good predictor of their behavior under a national system, but it does provide an excellent subgroup to test whether *any* agency types might be expected to make significant changes in their transfer patterns under a nationwide program.

## B. ESTIMATED IMPACTS AMONG SUBGROUPS

Among for-profit and nonprofit agencies, a small number of the impacts on selection and retention were significant, but none was in the direction to suggest reduced access for higher-cost patients (Table V.1). With respect to the measures of patient selection, for example, the estimated impact on expected service utilization was positive for both subgroups and large enough in the case of for-profit agencies to be marginally significant. For-profit agencies also experienced marginally significant reductions in their rate of transfer under both measures examined, whereas nonprofits showed an insignificant increase.

Hospital-based agencies experienced a highly significant increase in the predicted service use of their patients under the demonstration (Table V.2). As with for-profits, this impact was not consistent with the incentive to select a lower-cost mix, and certainly not evidence that these agencies tried to avoid admitting higher-cost patients. For the remaining measures, the impacts for hospital-based agencies were sometimes large but not significant or in the direction hypothesized. All impacts among freestanding agencies were insignificant and generally small.

A probable explanation for the highly significant selection impact among hospital-based agencies is that the very small subgroup sample makes the impact estimates for this subgroup sensitive to random events. In particular, with only six hospital-based agencies in the treatment group and four in the control group, any impact estimate might be highly affected by how these agencies were randomly allocated to the treatment and control groups and how factors *unrelated to the demonstration* influenced the patient mix of each.

An alternative, though less likely, explanation for this selection impact is that the hospital-based agencies in the treatment group intentionally increased the number of patients in the higher-cost Home Health Utilization Groups (HHUGs). The possible incentives to pursue this strategy are

## TABLE V.1

## IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY IS FOR-PROFIT OR NONPROFIT

| | For-Profit Impact[c] (*p*-value)[d] | Nonprofit Impact[c] (*p*-value)[d] | Difference[c] (*p*-value)[d] |
|---|---|---|---|
| **Selection Measures[a]** | | | |
| Predicted Service Utilization | 187*<br>(.06) | 31<br>(.77) | 156<br>(.30) |
| Has Diabetes | -1.8<br>(.26) | 4.4**<br>(.01) | -6.2**<br>(.01) |
| Has Serious Decubiti | 0.7<br>(.50) | -0.4<br>(.72) | 1.1<br>(.57) |
| Has Paid Help Available | 0.7<br>(.67) | 0.4<br>(.67) | 0.3<br>(.98) |
| **Retention Measures[b]** | | | |
| Transfer by One Year Postadmission, Base Measure (Percentage) | -0.6*<br>(.07) | 0.2<br>(.90) | -0.8<br>(.23) |
| Transfer by One Year Postadmission, Less Restrictive Measure (Percentage) | -0.7*<br>(.09) | 0.8<br>(.23) | -1.5*<br>(.06) |
| **Sample Size for Selection Measures** | | | |
| **Observations** | 13,317 | 41,036 | |
| **Agencies** | 42 | 42 | |
| **Sample Size for Retention Measures** | | | |
| **Observations** | 12,340 | 39,793 | |
| **Agencies** | 40 | 39 | |

SOURCE: Selection measures obtained from Medicare claims, QA admission forms, UB-92 remarks fields, Medicare enrollment files, and agency cost reports. Retention measures obtained from Medicare claims data.

NOTE: Each observation in the sample reflects a patient's admission to (episode start in) a home health agency participating in the demonstration. The selection sample includes all available episodes. The retention sample includes only initial episodes in the first demonstration year.

[a] Measured at home health admission. Predicted utilization measure is based on the primary measure (which excludes planned treatments) and reflects service use during the first 120 days after admission (the at-risk period). Serious decubiti are diagnosed as stage 3 or 4.

[b] Base transfer measure: 0 to 3 days between discharge and other-agency admission, or 0 to 7 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date. Less restrictive measure: 0 to 7 days between discharge and other-agency admission, or 0 to 14 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date.

[c] Regression-adjusted to control for preexisting differences between treatment and control agencies.

[d] The *p*-value corresponds to a test of whether the corresponding impact is statistically different from zero. It is based on standard errors inflated to account for the effects of clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

## TABLE V.2

## IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY IS HOSPITAL-BASED OR FREESTANDING

| | Hospital-Based Impact[c] (*p*-value)[d] | Freestanding Impact[c] (*p*-value)[d] | Difference[c] (*p*-value)[d] |
|---|---|---|---|
| **Selection Measures**[a] | | | |
| Predicted Service Utilization | 696***<br>(.00) | 30<br>(.68) | 666***<br>(.00) |
| Has Diabetes | 7.2<br>(.12) | 0.4<br>(.63) | 6.8<br>(.17) |
| Has Serious Decubiti | 1.1<br>(.38) | 0.1<br>(.88) | 1.0<br>(.43) |
| Has Paid Help Available | 3.7<br>(.40) | 0.3<br>(.70) | 3.4<br>(.47) |
| **Retention Measures**[b] | | | |
| Transfer by One Year Postadmission, Base Measure (Percentage) | -0.5<br>(.53) | -0.2<br>(.25) | -0.3<br>(.74) |
| Transfer by One Year Postadmission, Less Restrictive Measure (Percentage) | -1.0<br>(.36) | 0.1<br>(.97) | -1.1<br>(.40) |
| **Sample Size for Selection Measures** | | | |
| **Observations** | **6,865** | **47,488** | |
| **Agencies** | **10** | **74** | |
| **Sample Size for Retention Measures** | | | |
| **Observations** | **5,285** | **46,848** | |
| **Agencies** | **9** | **70** | |

SOURCE: Selection measures obtained from Medicare claims, QA admission forms, UB-92 remarks fields, Medicare enrollment files, and agency cost reports. Retention measures obtained from Medicare claims data.

NOTE: Each observation in the sample reflects a patient's admission to (episode start in) a home health agency participating in the demonstration. The selection sample includes all available episodes. The retention sample includes only initial episodes in the first demonstration year.

[a] Measured at home health admission. Predicted utilization measure is based on the primary measure (which excludes planned treatments) and reflects service use during the first 120 days after admission (the at-risk period). Serious decubiti are diagnosed as stage 3 or 4.

[b] Base transfer measure: 0 to 3 days between discharge and other-agency admission, or 0 to 7 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date. Less restrictive measure: 0 to 7 days between discharge and other-agency admission, or 0 to 14 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date.

[c] Regression-adjusted to control for preexisting differences between treatment and control agencies.

[d] The *p*-value shown in parentheses corresponds to a test of whether the corresponding impact is statistically different from zero. It is based on standard errors inflated to account for the effects of clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

twofold. First, if the agency could obtain relatively low utilizers of care within high-cost HHUGs, its (case-mix-adjusted) payments might rise by more than its service costs *even with no change in its care provision practices*. Second, agencies might have found more opportunity for large-scale cutbacks in services to patients in the higher-cost HHUGs, making it easier to bring their episode costs below the (upwardly adjusted) payment. For most agency types, this strategy was far less likely to be adopted than one to avoid high-need patients because it was both more complex and obscure. However, hospital-based agencies might have found this strategy somewhat more tenable as they had access to detailed information on most potential admissions; moreover, they could have potentially raised the proportion of patients into higher-cost HHUGs simply by admitting a large share of patients to the hospital for an intervening (within-episode) stay.

In results not shown, we did find that hospital-based agencies in the treatment group experienced a significant increase (close to seven percentage points) between the predemonstration and demonstration in the proportion of their patients in a higher-cost HHUG. However, when we controlled for this increase in an additional specification, the expected patient utilization among hospital-based agencies in the treatment group remained positive, large, and significant relative to those in the control group.[2] Moreover, when estimating impacts within selected high-cost HHUGs (those with sufficient sample sizes), we found that the expected costs of patients in hospital-based treatment group agencies were equal to or greater than those in hospital-based control agencies. Thus, it appears that hospital-based agencies did experience a worsening of their patient mix under the demonstration but that this change was not attributable to the system of payment.

---

[2] This specification included a set of indicator (control) variables that accounted for each patient's HHUG.

The effects on selection among small agencies and low-use practice patterns (Tables V.3 and V.4, respectively) were similar to those for hospital-based agencies. Both types of agency displayed a significant increase in their patients' expected utilization and in their share of patients with diabetes, neither of which was consistent with the incentives to obtain a more favorable patient mix. Meanwhile, the estimated impacts among their respective counterparts, large agencies and high-use agencies, were generally small and insignificant. The only effects of note were a modest but significant decline in the rate of transfer under the base measure, which, again, would not suggest efforts to obtain a lower-cost mix.

The significant growth in expected utilization among small and, in particular, low-use treatment group agencies is more difficult to attribute to random events since their sample sizes are considerably larger than hospital-based agencies. Like hospital-based agencies, these subgroups did experience growth in the proportion of their patients in higher-cost case-mix groups. When we controlled for these changes in case mix, however, the expected costs of patients in small and low-use treatment group agencies remained significantly larger than their control group counterparts. These results thus contradict the hypothesis that these two types of agency intentionally admitted higher-cost patients in order to more easily bring their costs below the case-mix-adjusted payment. Instead, like hospital-based agencies, it appears that these two subgroups experienced adverse patient selection under the demonstration attributable to random events rather than the system of payment.

In the final set of results, shown in Table V.5, we find no evidence that chain and nonchain agencies avoided high utilizers of care. Like several other subgroups, chain member agencies in the treatment group admitted a higher proportion of patients with diabetes than those in the control group. However, this difference was more isolated than for previous agency subgroups, and thus was less indicative of a significant change in the patient mix of chain agencies.

## TABLE V.3

## IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY IS SMALL OR LARGE

| | Small Agency Impact[c] (*p*-value)[d] | Large Agency Impact[c] (*p*-value)[d] | Difference[c] (*p*-value)[d] |
|---|---|---|---|
| **Selection Measures[a]** | | | |
| Predicted Service Utilization | 303**<br>(.03) | 40<br>(.60) | 263*<br>(.10) |
| Has Diabetes | 6.3**<br>(.03) | -0.8<br>(.65) | 7.2**<br>(.03) |
| Has Serious Decubiti | 2.2<br>(.17) | -0.4<br>(.53) | 2.6<br>(.13) |
| Has Paid Help Available | 1.9<br>(.46) | 0.1<br>(.71) | 1.8<br>(.57) |
| **Retention Measures[b]** | | | |
| Transfer by One Year Postadmission, Base Measure (Percentage) | 0.7<br>(.30) | -0.7**<br>(.02) | 1.4**<br>(.04) |
| Transfer by One Year Postadmission, Less Restrictive Measure (Percentage) | 1.1<br>(.20) | -0.6<br>(.15) | 1.7*<br>(.08) |
| **Sample Size for Selection Measures** | | | |
| **Observations** | **3,396** | **50,957** | |
| **Agencies** | **22** | **62** | |
| **Sample Size for Retention Measures** | | | |
| **Observations** | **2,781** | **49,352** | |
| **Agencies** | **22** | **57** | |

SOURCE: Selection measures obtained from Medicare claims, QA admission forms, UB-92 remarks fields, Medicare enrollment files, and agency cost reports. Retention measures obtained from Medicare claims data.

NOTE: Each observation in the sample reflects a patient's admission to (episode start in) a home health agency participating in the demonstration. The selection sample includes all available episodes. The retention sample includes only initial episodes in the first demonstration year. Large (small) agencies are defined as those that provided more (fewer) than 30,000 visits during the predemonstration year.

[a] Measured at home health admission. Predicted utilization measure is based on the primary measure (which excludes planned treatments) and reflects service use during the first 120 days after admission (the at-risk period). Serious decubiti are diagnosed as stage 3 or 4.

[b] Base transfer measure: 0 to 3 days between discharge and other-agency admission, or 0 to 7 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date. Less restrictive measure: 0 to 7 days between discharge and other-agency admission, or 0 to 14 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date.

[c] Regression-adjusted to control for preexisting differences between treatment and control agencies.

[d] The *p*-value shown in parentheses corresponds to a test of whether the corresponding impact is statistically different from zero. It is based on standard errors inflated to account for the effects of clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

## TABLE V.4

## IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY HAS A HIGH-USE OR LOW-USE PRACTICE PATTERN

| | High-Use Practice Pattern Impact[c] (*p*-value)[d] | Low-Use Practice Pattern Impact[c] (*p*-value)[d] | Difference[c] (*p*-value)[d] |
|---|---|---|---|
| **Selection Measures[a]** | | | |
| Predicted Service Utilization | -52<br>(.62) | 270***<br>(.00) | -322**<br>(.03) |
| Has Diabetes | -1.3<br>(.37) | 3.7***<br>(.01) | -5.0***<br>(.01) |
| Has Serious Decubiti | 0.3<br>(.92) | 0.1<br>(.99) | 0.2<br>(.95) |
| Has Paid Help Available | 3.5**<br>(.04) | -2.5<br>(.37) | 6.0*<br>(.07) |
| **Retention Measures[b]** | | | |
| Transfer by One Year Postadmission, Base Measure (Percentage) | -0.6**<br>(.02) | 0.5<br>(.67) | -1.1*<br>(.07) |
| Transfer by One Year Postadmission, Less Restrictive Measure (Percentage) | -0.3<br>(.57) | -0.6<br>(.24) | -0.9<br>(.25) |
| **Sample Size for Selection Measures** | | | |
| **Observations** | **26,316** | **28,037** | |
| **Agencies** | **42** | **42** | |
| **Sample Size for Retention Measures** | | | |
| **Observations** | **23,907** | **28,226** | |
| **Agencies** | **42** | **37** | |

SOURCE: Selection measures obtained from Medicare claims, QA admission forms, UB-92 remarks fields, Medicare enrollment files, and agency cost reports. Retention measures obtained from Medicare claims data.

NOTE: Each observation in the sample reflects a patient's admission to (episode start in) a home health agency participating in the demonstration. The selection sample includes all available episodes. The retention sample includes only initial episodes in the first demonstration year. An agency is defined as having a high-use practice pattern if its (case-mix-adjusted) number of visits per episode in the base (predemonstration) period was above the median for all agencies in the sample. Otherwise, it is defined as having a low-use practice pattern.

[a] Measured at home health admission. Predicted utilization measure is based on the primary measure (which excludes planned treatments) and reflects service use during the first 120 days after admission (the at-risk period). Serious decubiti are diagnosed as stage 3 or 4.

[b] Base transfer measure: 0 to 3 days between discharge and other-agency admission, or 0 to 7 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date. Less restrictive measure: 0 to 7 days between discharge and other-agency admission, or 0 to 14 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date.

[c] Regression-adjusted to control for preexisting differences between treatment and control agencies.

[d] The *p*-value shown in parentheses corresponds to a test of whether the corresponding impact is statistically different from zero. It is based on standard errors inflated to account for the effects of clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

## TABLE V.5

## IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY IS A CHAIN MEMBER

| | Chain Member Impact[c] (*p*-value)[d] | Not Chain Member Impact[c] (*p*-value)[d] | Difference[c] (*p*-value)[d] |
|---|---|---|---|
| **Selection Measures** | | | |
| Predicted Service Utilization | 165<br>(.11) | 82<br>(.36) | 83<br>(.56) |
| Has Diabetes | 4.4**<br>(.01) | -0.5<br>(.87) | 4.9**<br>(.03) |
| Has Serious Decubiti Condition | 1.1<br>(.46) | -0.3<br>(.63) | 1.4<br>(.41) |
| Has Paid Help Available | -0.4<br>(.98) | 1.1<br>(.41) | -1.5<br>(.65) |
| **Retention Measures** | | | |
| Transfer by One Year Postadmission, Base Measure (Percentage) | 0.4<br>(.58) | -0.7*<br>(.05) | 1.1<br>(.14) |
| Transfer by One Year Postadmission, Less Restrictive Measure (Percentage) | 0.8<br>(.25) | -0.7<br>(.18) | 1.5<br>(.13) |
| Sample Size for **Selection Measures** | | | |
| **Observations** | **12,846** | **41,287** | |
| **Agencies** | **27** | **57** | |
| Sample Size for **Retention Measures** | | | |
| **Observations** | **13,177** | **39,016** | |
| **Agencies** | **28** | **51** | |

SOURCE: Selection measures obtained from Medicare claims, QA admission forms, UB-92 remarks fields, Medicare enrollment files, and agency cost reports. Retention measures obtained from Medicare claims data.

NOTE: Each observation in the sample reflects a patient's admission to (episode start in) a home health agency. The selection sample includes all available episodes. The retention sample includes only initial episodes in the first demonstration year.

[a] Measured at home health admission. Predicted utilization measure is based on the primary measure (which excludes planned treatments) and reflects service use during the first 120 days after admission (the at-risk period). Serious decubiti are diagnosed as stage 3 or 4.

[b] Base transfer measure: 0 to 3 days between discharge and other-agency admission, or 0 to 7 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date. Less restrictive measure: 0 to 7 days between discharge and other-agency admission, or 0 to 14 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date.

[c] Regression-adjusted to control for preexisting differences between treatment and control agencies.

[d] The *p*-value shown in parentheses corresponds to a test of whether the corresponding impact is statistically different from zero. It is based on standard errors inflated to account for the effects of clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

## C. SUMMARY

Impact estimates for selected agency subgroups underscore the main finding--that the demonstration had no adverse affects on access to care for higher-cost patients. For each subgroup investigated, most of the estimated impacts on patient selection or retention were insignificant. Moreover, among the set of significant impacts, each suggested that the demonstration had actually raised selection of higher-cost patients or lowered the overall rate of transfer.

This set of significant impacts is difficult to attribute to the demonstration, because it completely contradicts treatment agencies' incentive to obtain and keep only relatively low-needs patients. In fact, the subgroups in which most of these significant treatment-control differences occurred (for-profit agencies, hospital-based agencies, low-use agencies) were those hypothesized as most likely to pursue a strategy of favorable selection under the demonstration. In an additional analysis, we investigated the alternative hypothesis that these agency subgroups might actually have pursued higher-cost patients under the demonstration, presumably in an effort to drive their case-mix-adjusted payment above their service costs. We find no evidence to support this hypothesis. Instead, after controlling for the actual case-mix category of each patient, we find that treatment group agencies still admitted a patient mix with significantly higher expected utilization than their control group counterparts. Thus, it appears that these subgroups simply experienced a higher-cost "draw" of patients under the demonstration, a change that was unrelated to the system of payment.

# VI. CONCLUSIONS

Based on a thorough analysis of data from the first two years of the Per-Episode Home Health Prospective Payment Demonstration, we find no evidence that prospective payment changed the selection or retention practices of participating agencies. The consistency and robustness of this finding suggest that it can serve as a useful guide to policymakers seeking to implement some form of prospective payment system as mandated by the Balanced Budget Act of 1997.

## A. KEY FINDINGS

Per-episode prospective payment (as defined by the rules of the demonstration) did not lead agencies to avoid patients expected to be high utilizers of care. Treatment agencies were no less likely to admit a patient with a serious medical condition, limitations in activities of daily living (ADLs), or other conditions predictive of higher-than-average service needs. In addition, based on their observed characteristics at the start of an episode, the average patient in a treatment agency was predicted to use almost exactly the same level of services during the at-risk period as the average patient in a control agency.

Among the roughly 1 in 10 patients with multiple episodes of care in the same agency, those readmitted to treatment agencies were found to have slightly lower predicted service use (on average) than those readmitted to control agencies. While consistent with the incentives to select a lower-cost patient mix under the demonstration, this difference appears to be due to difficulties measuring expected service use among the treatment group rather than to a true change in selection practices. The expected service use measure depends in part on prior home health utilization, which our other work has shown to be significantly lower for treatment agency patients. Nevertheless, we cannot reject the possibility that a small degree of favorable selection took place among readmissions to the treatment group, particularly since they are such an attractive target for such efforts.

We also find no evidence that the demonstration affected the admission of patients expected to have relatively high costs per visit. For example, the demonstration did not change the proportion of patients admitted from isolated rural areas or from areas with a high percentage of non-English-speaking households. Similarly, several additional measures of rurality and language isolation showed no effects from the demonstration.

Agencies also displayed no greater propensity to transfer high-use patients after they were admitted. In each month after admission, the rates of transfer in both treatment and control agencies were nearly identical and quite low. By one year after admission, for example, 3.2 percent of patients had been transferred from control agencies, compared to 2.9 percent from treatment agencies (regression-adjusted).

Impact estimates for selected agency subgroups underscore the main finding that the demonstration had no adverse affects on access to care for higher-cost patients. For the 10 subgroups investigated, most of the estimated impacts on patient selection or retention were insignificant. Moreover, among the set of significant impacts, each suggested that the demonstration actually had raised selection of high utilizers of care or had lowered the rate of transfer. Most of these significant impacts were associated with three agency types--hospital-based agencies, agencies with low-use practice patterns, and small agencies. After evaluating whether these impacts were truly due to differences in the system of payment, we conclude, instead, that these three subgroups simply experienced an adverse patient draw that was unrelated to per-episode payment.

All the findings described above proved highly robust. Impact estimates showed essentially no change over the course of the demonstration, and they were not affected by the use of sample weights or by outlier values. We also find no evidence that our estimates were affected by the loss of observations from our samples due to missing data.

## B. LIMITATIONS OF THE ANALYSIS

Despite the robustness of our findings, this study does have some limitations. Perhaps the most important potential limitation is the extent to which we can generalize our findings to home health agencies nationwide. Like any study design based on voluntary participation, the agencies in this demonstration may reflect a group quite different from agencies that did not participate. For example, if the agencies in the demonstration were those most capable of cutting service use in response to the demonstration incentives, they might have had less reason to pursue alternative strategies to lower per-episode costs. If true, our demonstration impacts might understate the actual changes in selection and retention that would take place nationally under a similar payment system. One key factor that suggests our results can be widely generalized, however, is that agencies of all types displayed no significant changes in their patterns of selection or retention in response to the demonstration. Although agencies of differing sizes, auspice, or practice patterns might be expected to differ, both in their capacity and their desire to respond to demonstration incentives, none of these agencies displayed an effort to admit lower-cost patients or to transfer high-cost ones more quickly or more often.

A second limitation, related to the issue of generalizability, is that a national program of prospective payment would almost certainly differ from the program implemented for the demonstration. For example, under a national system, agencies would not be protected from incurring financial losses, which could lead some agencies to respond more aggressively and to try to obtain a lower-cost patient mix. Moreover, an agency's case-mix-adjusted payment would almost certainly not be based solely on its own prior experiences. For some agencies, particularly those with relatively higher costs per episode, this might lead to substantial changes in patient selection or retention as they tried to bring their costs down below the payment.

A third limitation of this report is that the findings are based on agencies' behavior over a relatively short period of time. In particular, given the strong competition in the market for home health care, many agencies could probably only make large-scale changes in their referral sources over an extended period, perhaps years in length. This suggests that the period evaluated, the first two years of a three-year demonstration, might have been too short to identify material changes in agencies' patterns of selection. At the same time, it is important to note that we found just as little evidence of selection effects occurring during the second year of the demonstration as we did during the first year.

A final limitation is that this report focuses narrowly on one possible effect of the demonstration on patient care. Other reports will examine the effects of the demonstration on other aspects of patient care, including quality, other Medicare service use, and the use of informal care and services paid by funders other than Medicare.[1] In addition, a recent report by Trenholm (2000) examined impacts on the use of services; a later report will examine the impacts of the demonstration on agency cost per visit and cost per episode, as well as on profit and loss. A final report will combine the findings across outcome measures.

---

[1] Preliminary reports on the effect of per-episode payment on care quality (Chen 2000) and the use of other Medicare services (Schore 2000) suggested that the demonstration had no effect on these outcomes during the first demonstration year.

## C. POLICY IMPLICATIONS

The findings from this report provide key information on the effects of prospective payment at the very time when HCFA, under mandate from the 1997 Balanced Budget Act, must design and implement a nationwide prospective payment system. The key result from this study--that agencies did not significantly alter their selection or retention practices in response to per-episode prospective payment--is highly relevant to HCFA's efforts. When combined with the findings of large-scale reductions in service use under the demonstration (Trenholm 2000), this result suggests that agencies responded to the per-episode prospective payment in precisely the way HCFA anticipated. It thus underscores the potential for prospective payment to be a workable payment system for the Medicare home health benefit.

One additional result that is relevant to the design of a national payment system is the adverse selection experienced by certain agencies under the demonstration. This result was not due to changes in agencies' selection practices; in fact, it appears to be entirely unrelated to the demonstration. Thus, it underscores the need for an effective case-mix adjuster under a nationwide system of per-episode payment, since, without an effective adjuster, the random variation in patient mix over time might place many agencies in financial trouble. For example, our subgroup results show that small treatment agencies experienced a significant increase of about nine percent in the expected cost of their patients, compared to their control group counterparts. While the demonstration case-mix adjuster accounted for some of this increase (roughly one-third), this adverse selection might still have been enough to threaten the solvency of at least some of these agencies if they lacked the loss protection and other features of the demonstration to protect them.

# REFERENCES

Brown, Randall, Barbara Phillips, Christine Bishop, Amy Klein, Grant Ritter, Craig Thornton, Peter Schochet, and Kathleen Skwara. "The Effects of Predetermined Payment Rates for Home Health Care." Princeton, NJ: Mathematica Policy Research, Inc., December 1995.

Chen, Arnold. "The Impacts of Per-Episode Prospective Payment for Medicare Home Health Care on the Quality of Care: Less Is Not Necessarily Worse." Princeton, NJ: Mathematica Policy Research, Inc., May 9, 2000.

Health Care Financing Review. *Medicare and Medicaid Statistical Supplement, 1996.*

*Health Care Financing Review. Statistical Supplement*, 1998.

Phillips, Barbara R., Randall Brown, Jennifer Schore, Amy C. Klein, Peter Z. Schochet, Jerrold W. Hill, and Craig Thornton. "Evaluation of the Per-Episode Home Health Prospective Payment Demonstration Design." Report submitted to the Health Care Financing Administration. Princeton, NJ: Mathematica Policy Research, Inc., September 1995.

Schore, Jennifer. "The Impact of Home Health Prospective Payment on Medicare Service Use and Reimbursement." Princeton, NJ: Mathematica Policy Research, Inc., March 2000.

Teplitsky, Sanford, V. and Mary Ann Janson (eds.). *Home Health & Hospice Manual: Regulations and Guidelines*. Owings Mills, MD: National Health Publishing, 1985-1992, pp. VII.23, Section 204.4.

Trenholm, Christopher. "Per-Episode Prospective Payment for Medicare Home Health Care Sharply Reduces Service Use." Princeton, NJ: Mathematica Policy Research, Inc., May 10, 2000.

U.S. General Accounting Office. *Medicare Health Benefit: Impact of Interim Payment System and Agency Closures on Access to Services.* Washington, DC: GAO, September 1998.

U.S. General Accounting Office. *Medicare Home Health Care. Prospective Payment System Will Need Refinement as Data Becomes Available.* Washington, DC: GAO, April 2000.

# APPENDIX A

# SUMMARY OF CASE-MIX ADJUSTMENT FOR PAYMENTS DURING THE DEMONSTRATION

During the demonstration, agency payment rates were adjusted for severity, using the Home Health Utilization Groups (HHUGs) Classification system. This system, which was specifically developed for the demonstration, classified patients into 1 of 18 mutually exclusive cells on the basis of the following information (see Table A.1):

- Whether the patient had an intervening hospital stay in an acute-care hospital during the 120 days following a home health admission
- Whether the patient was impaired in at least four of five activities of daily living (bathing, eating/tube feeding, dressing, toileting/elimination, and transferring) at home health admission
- Whether wound care was planned for the patient at home health admission
- Whether the patient was discharged from a hospital within the 14 days preceding home health admission
- Whether the patient had stage 3 or stage 4 decubitus ulcer at home health admission
- Whether the patient had cancer at home health admission that affected current treatment or personal-care needs
- Whether the patient has had a cerebrovascular accident that affected current treatment or personal-care needs
- Whether the patient had diabetes that affected current treatment or personal-care needs

As with per-episode payment rate, demonstration case-mix adjustment was based on the agency's base-period experience. Information on the characteristics listed above was collected on each patient admitted to each agency in the 90 days ("base quarter") before it began to operate under the demonstration. The same information was collected throughout the demonstration in the "remarks" section of the Uniform Bill, HCFA form 92 (UB-92).

TABLE A.1

## NATIONAL HOME HEALTH AGENCY PROSPECTIVE PAYMENT DEMONSTRATION PATIENT CLASSIFICATION SYSTEM FOR CASE-MIX ADJUSTMENT

| | ADMISSION CHARACTERISTICS | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Hospital stay during 120 days? (from HCFA claims data) | ADL limitations (of 5) | Need wound care? | Preadmission location is a hospital? | Decubitus stage 3 or 4? | Diabetes? | CVA? | Cancer? | CASE-MIX Group: | Average Patient Cost in the Base Quarter[a] (Dollars) |
| YES | 0-3 | YES | YES | →→→→→→→ | →→→→→→→ | →→→ | →→→ | =1 | $5,871 |
| YES | 0-3 | YES | NO | →→→→→→→ | →→→→→→→ | →→→ | →→→ | =2 | 5,190 |
| YES | 0-3 | NO | YES | →→→→→→→ | →→→→→→→ | →→→ | →→→ | =3 | 3,159 |
| YES | 0-3 | NO | NO | →→→→→→→ | →→→→→→→ | →→→ | →→→ | =4 | 3,197 |
| YES | 4-5 | YES | →→→→→→→ | → YES | →→→→→→→ | →→→ | →→→ | =5 | 6,539 |
| YES | 4-5 | YES | →→→→→→→ | → NO | →→→→→→→ | →→→ | →→→ | =6 | 5,340 |
| YES | 4-5 | NO | →→→→→→→ | →→→→→→→ | →→→→→→→ | → YES | →→→ | =7 | 5,287 |
| YES | 4-5 | NO | →→→→→→→ | →→→→→→→ | →→→→→→→ | → NO | YES | =8 | 3,440 |
| YES | 4-5 | NO | →→→→→→→ | →→→→→→→ | →→→→→→→ | →NO | NO | =9 | 4,128 |
| NO | 0-1 | YES | YES | →→→→→→→ | →→→→→→→ | →→→ | →→→ | =10 | 3,813 |
| NO | 0-1 | YES | NO | →→→→→→→ | →→→→→→→ | →→→ | →→→ | =11 | 3,933 |
| NO | 0-1 | NO | →→→→→→→ | →→→→→→→ | → YES | →→→ | →→→ | =12 | 2,977 |
| NO | 0-1 | NO | →→→→→→→ | →→→→→→→ | → NO | →→→ | →→→ | =13 | 2,674 |
| NO | 2-5 | YES | →→→→→→→ | → YES | →→→→→→→ | →→→ | →→→ | =14 | 8,041 |
| NO | 2-5 | YES | →→→→→→→ | → NO | →→→→→→→ | →→→ | →→→ | =15 | 5,263 |
| NO | 2-5 | NO | →→→→→→→ | →→→→→→→ | → YES | →→→ | →→→ | =16 | 4,888 |
| NO | 2-5 | NO | →→→→→→→ | →→→→→→→ | → NO | YES | →→→ | =17 | 4,919 |
| NO | 2-5 | NO | →→→→→→→ | →→→→→→→ | → NO | NO | →→→ | =18 | 3,488 |

[a]Calculated by multiplying (for each patient in a given HHUG) the number of visits of each type during the first 120 days postadmission by the respective cost-per-visit limit.

In order to determine the case-mix adjustment during the demonstration for each agency, Abt Associates first calculated a base quarter index for each agency. To do this, Abt calculated a category weight for each of the 18 case-mix groups by dividing the average cost in each group by the agency's overall average cost.[1] Abt then multiplied each category weight by the percentage of episodes that fell into each case-mix category during the base quarter and summed across all 18 categories. The result was a base-quarter index for each agency, which was always equal to one. Abt calculated a similar index for the demonstration years by multiplying the category weight from the base quarter times the percentage of episodes that fell into each case-mix category during the relevant year, then summing across all 18 categories. The demonstration year index for each agency was then divided by the agency's base-quarter index and multiplied by the agency's base-year episode rate. This resulted in the case-mix-adjusted episode rate for each agency.

As Table A.1 shows, the average cost of patients in the various HHUGs varied considerably in the "base" quarter prior to the demonstration start.[2] For example, patients classified in the lowest-cost case-mix category (HHUG 13) averaged $2,674 across agencies, while those classified in the highest-cost categories (HHUGs 5 and 14) averaged $6,539 and $8,041, respectively.

Given this variation, we created three basic, patient-level variables to measure agencies' case-mix distribution in the predemonstration and demonstration periods. The first variable, reflecting low-use patients, equaled one for patients classified in HHUG 13; zero otherwise. The second, for medium-use patients, equaled one for patients classified in HHUGs 3, 4, 8 to 12, and 18; zero otherwise. The third variable, for high-use patients, equaled one for patients classified in HHUGs 1, 2, 5 to 7, and 14 to 17; zero otherwise. We used these variables to control for preexisting differences in agencies' case-mix distributions, as well as to investigate changes in case mix under the demonstration.

---

[1] If an agency had no observations in a particular category during the base quarter, the average cost for that case-mix group calculated across all demonstration agencies was used instead of an agency-specific cost.

[2] These costs have been standardized across agencies by using the cost limit for each service rather than each agency's actual per-visit cost. They therefore differ considerably from the actual costs of patients in these HHUGs.

# APPENDIX B

## SUMMARY OF THE VARIABLE USED TO ACCOUNT FOR PREEXISTING DIFFERENCES IN AGENCY PRACTICE PATTERNS

The agency practice pattern variable is an index of the average number of visits per episode provided by an agency in the first 120 days of base-quarter episodes, relative to the average number provided by other demonstration agencies. The variable was used to control for preexisting differences across agencies in the average number of visits provided to patients, as well as to construct subgroups for agencies with relatively low- and high-use practice patterns. A value greater than one indicates that, controlling for variation in case mix, an agency provided more visits during the 120-day period than did other demonstration agencies during the quarter preceding the demonstration.

The specification of the variable is as follows. Let subscript *i* refer to the service type, and subscript *j* refer to the case-mix cell. Using the case-mix adjuster (developed by Abt Associates), we classify an agency's patients into 1 of 18 case-mix cells ($j$=1,..,18). Within each case-mix cell, we multiply the average visits of each type for a given agency ($n_{ij}$) by its national cost limit in the base year ($w_i$), then sum across the visit types. This sum essentially reflects a weighted count of the average visits for an agency within a case-mix cell. Second, we use these weighted counts to construct the ratio of the agency's average number of visits received by patients in the case-mix cell to the average among all agencies for this cell ($\Sigma w_i N_{ij}$). Finally, for each agency, we arrive at the practice pattern index by summing across the 18 case-mix ratios, weighting each ratio by the agency's proportion of episodes in the case-mix cell during the base quarter ($p_j$). Thus, for each agency, the index practice pattern is given by:

$$\sum_j p_j \left( \frac{\sum_i w_i n_{ij}}{\sum_i w_i N_{ij}} \right)$$

# APPENDIX C

# VARIABLE CONSTRUCTION FOR THE ANALYSIS OF SELECTION BASED ON EXPECTED SERVICE USE

**Constructing the Multivariate (Index) Measures**. We created these variables as follows. First, for each episode in a control agency, we measured the actual utilization level by multiplying the number of visits provided of each type by its per-visit cost and then summing across all visit types.[1] Next, using this control-group subsample, we estimated a pair of regression models where the dependent variable was the utilization measure and the control variables were two sets of patient-level characteristics at admission.[2] The first regression, which excluded indicators of planned treatments, was used to create our so-called "primary" measure, on which we based most of our findings. (Results from this regression are shown in Table C.1.) The second regression, which included these indicators, was used to create our "secondary" measure.

As the last step, for every observation (episode) in our sample, we multiplied the value of each coefficient from the respective regressions by the value of the corresponding control variable, and then summed across all the control variables. For each patient, the resulting variables measured the anticipated utilization during the first 120 days post-admission, weighted by the per-unit cost of each service.

**Univariate Measures of Expected Service Use.** Based on the results from the regression model used to develop our "primary" index measure of expected service, we identified 12 individual variables that agencies might have used to obtain a favorable patient. Each of these variables, shown in italics in Table C.1, was selected because it had a significant impact of at least 10 percent on the

---

[1] This may also be thought of as a measure of each episode's expected costs. However, in the context of selection, we prefer to think of it as a service-use measure (where each service has been weighted by its average per-unit cost) in order to distinguish area-level factors (such as rurality) that may affect expected costs but presumably not expected utilization.

[2] Note that we excluded one control agency from this regression because it had substantial delays in processing claims that may have led to undercounts of actual service provision.

## TABLE C.1

## IMPACTS OF SELECTED CHARACTERISTICS OF CONTROL AGENCY PATIENTS AT ADMISSION ON THEIR HOME HEALTH UTILIZATION DURING THE SUBSEQUENT 120 DAYS

| Patient Characteristic[a] | Impact in Dollars ($p$-value)[b] |
|---|---|
| **Demographic Measures** | |
| Age Was Original Reason for Medicare Eligibility | -300<br>(.11) |
| Age: <65 Years | -116<br>(.65) |
| Age: 75 to 84 Years | 275***<br>(.00) |
| Age: >84 Years | 231<br>(.04) |
| Race/Ethnicity: White | -238<br>(.19) |
| Female | 94<br>(.20) |
| Enrolled in Medicaid | -51<br>(.70) |
| **Medical Conditions (for Case-Mix Adjustment)[c]** | |
| Cancer | -498**<br>(.01) |
| *Diabetes* | *748****<br>*(.00)* |
| *Cerebrovascular Accident (Stroke)* | *527****<br>*(.00)* |
| *Decubitus Ulcer: Stage 3 or 4* | *608***<br>*(.03)* |
| *Needs Complex Wound Care* | *334**<br>*(.10)* |
| Any Medical Condition (listed above) | 147<br>(.43) |

TABLE C.1 *(continued)*

| Patient Characteristic[a] | Impact in Dollars (*p*-value)[b] |
|---|---|
| **Other Conditions and Symptoms** | |
| Health Risk Factors (heavy smoking, obesity, alcoholism, and/or drug dependency) | -134<br>(.46) |
| Medical Regimen Changed in Last 14 Days | -17<br>(.89) |
| *Unstable at Admission* | *346****<br>*(.00)* |
| Urinary Tract Infection | 126<br>(.48) |
| *Urinary Incontinence* | *376****<br>*(.00)* |
| Dyspnea | 70<br>(.53) |
| Is Often Confused | 130<br>(.47) |
| Has Depressed Feelings | 66<br>(.56) |
| Displays Depressive Behaviors | -140<br>(.30) |
| Demonstrates Disruptive Behaviors | -315***<br>(.00) |
| Needs Assistance with Oral Medications | 246**<br>(.01) |
| *Surgical Wound Not Healing* | *622****<br>*(.05)* |
| **Limitations in ADLs (for Case-Mix Adjustment)[c]** | |
| Bathing | 244**<br>(.04) |
| Eating | 145<br>(.23) |
| Dressing | 317**<br>(.01) |
| Toileting | 83<br>(.49) |
| Transferring | 99<br>(.40) |
| Any ADL (listed above) | 239<br>(.16) |

TABLE C.1 *(continued)*

| Patient Characteristic[a] | Impact in Dollars ($p$-value)[b] |
|---|---|
| **Other Limitations in ADLs** | |
| *Ambulation* | *548****<br>*(.00)* |
| Prepare Light Meals | 50<br>(.62) |
| *Light Housekeeping* | *443****<br>*(.00)* |
| **Availability of Informal Care** | |
| *Paid Help Available* | *-353****<br>*(.00)* |
| Live-in Help Available | -156<br>(.04) |
| **Prognosis** | |
| Prognosis Is Good/Fair | 322<br>(.12) |
| *Life Expectancy Is More than 6 Months* | *434****<br>*(.01)* |
| Rehabilitative Prognosis is Poor | -33<br>(.81) |
| *Treatment Can Likely Be Taken Over*<br>*(by patient, friends, neighbors, or paid help)* | *-575****<br>*(.00)* |
| **Patient Prior Service Use Measures** | |
| Admitted to Home Health from Hospital | |
| Length of Hospital Stay Ending Within 14 Days Before Home Health Admission (Days) | -17<br>(.13) |
| Had SNF Stay Ending Within 14 Days Before Admission | 254***<br>(.01) |
| Total Part A (and Part B Inpatient/SNF/Home Health/Hospice) Reimbursement in Six Months Before Admission (Thousands) | 22***<br>(.00) |
| Home Health Visits in Six Months Before Admission | 7***<br>(.00) |

TABLE C.1 *(continued)*

| Patient Characteristic[a] | Impact in Dollars (*p*-value)[b] |
|---|---|
| **Other Patient-Level Variables (Percentage)** | |
| Had Medicare for Less than Six Months Before Admission | 431*<br>(.06) |
| Enrolled in HMO in Six Months Before Admission | 332<br>(.27) |
| Had Medicare as Second Payer in Six Months Before Admission | 77<br>(.74) |
| Admission Date: 4/1/96 Through 6/30/96 | 341<br>(.18) |
| Admission Date: 7/1/96 Through 9/30/96 | 131<br>(.50) |
| Admission Date: 10/1/96 Through 12/31/96 | 95<br>(.46) |
| Admission Date: 4/1/97 Through 6/30/97 | -221<br>(.10) |
| Admission Date: 7/1/97 Through 8/31/97 | -475**<br>(.04) |
| Demonstration Year 2 | 421*<br>(.06) |
| Demonstration Year 3 | 874*<br>(.10) |
| **Sample Size** | |
| **Observations** | **21,582** |
| **Agencies** | **38** |

SOURCE: Patient characteristics based on information from Medicare claims, QA admission forms, UB-92 remarks field (on the admission claim), and Medicare enrollment files. See Table II.1 for a complete breakdown of these variables by source.

NOTE: Each observation in the sample reflects a patient's episode start in a home health agency participating in the demonstration as part of the control group. The observations have been weighted to give agencies equal representation in the calculations.

[a] Variables shown in italics were also investigated individually for evidence of selection.

[b] The *p*-value corresponds to a test of whether the impact shown is statistically different from zero. It is based on standard errors inflated to account for the effects of clustering and weighting.

[c] These variables, which were obtained from the UB-92 remarks field on the admission claim form, were used to construct the case-mix adjuster used during the demonstration. For more information on this adjuster, see Appendix A.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

level of utilization relative to the mean value among the control group (about \$3,200)[3] and was therefore an attractive measure by which treatment agencies might pursue a lower-use patient mix.

---

[3] The indicator variable for cancer had a very large effect on the level of home health use, but it was in a negative direction. While reasonable explanations exist for this result (for example, home health episodes for cancer patients might often terminate quickly in a hospice or hospital admission), it seems unlikely that treatment agencies would have adopted a policy of selecting cancer patients in order to lower utilization.

# APPENDIX D

# DETAILED SUBGROUP TABLES

## TABLE D.1

## IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY IS FOR-PROFIT OR NONPROFIT

| | Treatment Group Mean[c] | Control Group Mean | Difference[c] |
|---|---|---|---|
| **Selection Measures[a]** | | | |
| Predicted Service Utilization | | | |
| For-profit | 3,610 | 3,423 | 187* |
| Nonprofit | 3,012 | 2,981 | 31 |
| Subgroup difference | | | 156 |
| Has Diabetes | | | |
| For-profit | 20.9 | 22.7 | -1.8 |
| Nonprofit | 21.7 | 17.3 | 4.4** |
| Subgroup difference | | | -6.2** |
| Has Serious Decubiti | | | |
| For-profit | 4.6 | 3.9 | 0.7 |
| Nonprofit | 2.8 | 3.2 | -0.4 |
| Subgroup difference | | | 1.1 |
| Has Paid Help Available | | | |
| For-profit | 18.9 | 18.2 | 0.7 |
| Nonprofit | 13.6 | 13.2 | 0.4 |
| Subgroup difference | | | 0.3 |
| **Retention Measures[b]** | | | |
| Transfer by One Year Postadmission, Base Measure (Percentage) | | | |
| For-profit | 3.3 | 3.9 | -0.6* |
| Nonprofit | 2.6 | 2.4 | 0.2 |
| Subgroup difference | | | -0.8 |
| Transfer by OneYear Postadmission, Less Restrictive Measure (Percentage) | | | |
| For-profit | 4.4 | 5.1 | -0.7* |
| Nonprofit | 3.9 | 3.1 | 0.8 |
| Subgroup difference | | | -1.5* |

| | **For-profit** | | **Nonprofit** | |
|---|---|---|---|---|
| **Sample Sizes** | **Selection** | **Retention** | **Selection** | **Retention** |
| **Treatment Group** | | | | |
| **Observations** | **6,666** | **5,422** | **25,452** | **20,962** |
| **Agencies** | **21** | **19** | **24** | **20** |
| **Control Group** | | | | |
| **Observations** | **6,651** | **6,918** | **15,584** | **18,831** |
| **Agencies** | **21** | **21** | **18** | **19** |

SOURCE: Selection measures obtained from Medicare claims, QA admission forms, UB-92 remarks fields, Medicare enrollment files, and agency cost reports. Retention measures obtained from Medicare claims data.

TABLE D.1 *(continued)*

---

NOTE: Each observation in the sample reflects a patient's admission to (episode start in) a home health agency participating in the demonstration. The selection sample includes all available episodes. The retention sample includes only initial episodes in the first demonstration year.

[a] Measured at home health admission. Predicted utilization measures is based on the primary measure (which excludes planned treatments) and reflects service use during the first 120 days after admission (the at-risk period). Serious decubiti are diagnosed as stage 3 or 4.

[b] Base transfer measure: 0 to 3 days between discharge and other-agency admission, or 0 to 7 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date. Less restrictive measure: 0 to 7 days between discharge and other-agency admission, or 0 to 14 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date.

[c] Regression-adjusted to control for preexisting differences between treatment and control agencies.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

## TABLE D.2

## IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY IS HOSPITAL-BASED OR FREESTANDING

| | Treatment Group Mean[c] | Control Group Mean | Difference[c] |
|---|---|---|---|
| **Selection Measures**[a] | | | |
| Predicted Service Utilization | | | |
| Hospital-based | 3,594 | 2,898 | 696*** |
| Freestanding | 3,308 | 3,278 | 30 |
| Subgroup difference | | | 666*** |
| Has Diabetes | | | |
| Hospital-based | 27.4 | 20.2 | 7.2 |
| Freestanding | 20.6 | 20.2 | 0.4 |
| Subgroup difference | | | 6.8 |
| Has Serious Decubiti | | | |
| Hospital-based | 3.8 | 2.7 | 1.1 |
| Freestanding | 3.8 | 3.7 | 0.1 |
| Subgroup difference | | | 1.0 |
| Has Paid Help Available | | | |
| Hospital-based | 22.8 | 19.1 | 3.7 |
| Freestanding | 15.5 | 15.2 | 0.3 |
| Subgroup difference | | | 3.4 |
| **Retention Measures**[b] | | | |
| Transfer by One Year Postadmission, Base Measure (Percentage) | | | |
| Hospital-based | 1.1 | 1.6 | -0.5 |
| Freestanding | 3.2 | 3.4 | -0.2 |
| Subgroup difference | | | -0.3 |
| Transfer by One Year Postadmission, Less Restrictive Measure (Percentage) | | | |
| Hospital-based | 1.4 | 2.4 | -1.0 |
| Freestanding | 4.6 | 4.5 | 0.1 |
| Subgroup difference | | | -1.1 |

| | **Hospital-Based** | | **Freestanding** | |
|---|---|---|---|---|
| **Sample Sizes** | **Selection** | **Retention** | **Selection** | **Retention** |
| **Treatment Group** | | | | |
| **Observations** | **2,043** | **868** | **30,075** | **25,516** |
| **Agencies** | **4** | **3** | **41** | **36** |
| **Control Group** | | | | |
| **Observations** | **4,822** | **4,417** | **17,413** | **21,332** |
| **Agencies** | **6** | **6** | **33** | **34** |

SOURCE: Selection measures obtained from Medicare claims, QA admission forms, UB-92 remarks fields, Medicare enrollment files, and agency cost reports. Retention measures obtained from Medicare claims data.

TABLE D.2 *(continued)*

---

NOTE: Each observation in the sample reflects a patient's admission to (episode start in) a home health agency participating in the demonstration. The selection sample includes all available episodes. The retention sample includes only initial episodes in the first demonstration year.

[a] Measured at home health admission. Predicted utilization measures is based on the primary measure (which excludes planned treatments) and reflects service use during the first 120 days after admission (the at-risk period). Serious decubiti are diagnosed as stage 3 or 4.

[b] Base transfer measure: 0 to 3 days between discharge and other-agency admission, or 0 to 7 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date. Less restrictive measure: 0 to 7 days between discharge and other-agency admission, or 0 to 14 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date.

[c] Regression-adjusted to control for preexisting differences between treatment and control agencies.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

## TABLE D.3

## IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY IS SMALL OR LARGE

| | Treatment Group Mean[c] | Control Group Mean | Difference[c] |
|---|---|---|---|
| **Selection Measures[a]** | | | |
| Predicted Service Utilization | | | |
| Small size | 3,675 | 3,372 | 303** |
| Large size | 3,226 | 3,186 | 40 |
| Subgroup difference | | | 263* |
| Has Diabetes | | | |
| Small size | 28.4 | 22.1 | 6.3* |
| Large size | 16.4 | 17.2 | -0.8 |
| Subgroup difference | | | 7.2** |
| Has Serious Decubiti | | | |
| Small size | 5.8 | 3.6 | 2.2 |
| Large size | 3.1 | 3.5 | -0.4 |
| Subgroup difference | | | 2.6 |
| Has Paid Help Available | | | |
| Small size | 16.2 | 14.3 | 1.9 |
| Large size | 18.6 | 18.5 | 0.1 |
| Subgroup difference | | | 1.8 |
| **Retention Measures[b]** | | | |
| Transfer by One Year Postadmission, Base Measure (Percentage) | | | |
| Small size | 5.5 | 4.8 | 0.7 |
| Large size | 2.0 | 2.7 | -0.7** |
| Subgroup difference | | | 1.4** |
| Transfer by One Year Postadmission, Less Restrictive Measure (Percentage) | | | |
| Small size | 7.0 | 5.9 | 1.1 |
| Large size | 3.1 | 3.7 | -0.6 |
| Subgroup difference | | | 1.7* |

| | **Small Size** | | **Large Size** | |
|---|---|---|---|---|
| **Sample Sizes** | **Selection** | **Retention** | **Selection** | **Retention** |
| **Treatment Group** | | | | |
| **Observations** | **2,724** | **2,013** | **29,394** | **24,371** |
| **Agencies** | **15** | **14** | **30** | **25** |
| **Control Group** | | | | |
| **Observations** | **672** | **768** | **21,563** | **24,981** |
| **Agencies** | **7** | **8** | **32** | **32** |

SOURCE: Selection measures obtained from Medicare claims, QA admission forms, UB-92 remarks fields, Medicare enrollment files, and agency cost reports. Retention measures obtained from Medicare claims data.

---

NOTE: Each observation in the sample reflects a patient's admission to (episode start in) a home health agency participating in the demonstration. The selection sample includes all available episodes. The retention sample includes only initial episodes in the first demonstration year. Large (small) agencies are defined as those that provided more (fewer) than 30,000 visits during the predemonstration year.

[a] Measured at home health admission. Predicted utilization measures is based on the primary measure (which excludes planned treatments) and reflects service use during the first 120 days after admission (the at-risk period). Serious decubiti are diagnosed as stage 3 or 4.

[b] Base transfer measure: 0 to 3 days between discharge and other-agency admission, or 0 to 7 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date. Less restrictive measure: 0 to 7 days between discharge and other-agency admission, or 0 to 14 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date.

[c] Regression-adjusted to control for preexisting differences between treatment and control agencies.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

## TABLE D.4

## IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY HAS A HIGH-USE OR LOW-USE PRACTICE PATTERN

| | Treatment Group Mean[c] | Control Group Mean | Difference[c] |
|---|---|---|---|
| **Selection Measures[a]** | | | |
| Predicted Service Utilization | | | |
| High-use practice pattern | 3,327 | 3,379 | -52 |
| Low-use practice pattern | 3,236 | 2,966 | 270*** |
| Subgroup difference | | | -322** |
| Has Diabetes | | | |
| High-use practice pattern | 18.0 | 19.3 | -1.3 |
| Low-use practice pattern | 24.0 | 20.3 | 3.7*** |
| Subgroup difference | | | -5.0*** |
| Has Serious Decubiti | | | |
| High-use practice pattern | 2.4 | 2.1 | 0.3 |
| Low-use practice pattern | 3.4 | 3.8 | 0.1 |
| Subgroup difference | | | 0.2 |
| Has Paid Help Available | | | |
| High-use practice pattern | 19.9 | 12.4 | 3.5 |
| Low-use practice pattern | 19.0 | 16.5 | -2.5 |
| Subgroup difference | | | 6.0* |
| **Retention Measures[b]** | | | |
| Transfer by One Year Postadmission, Base Measure (Percentage) | | | |
| High-use practice pattern | 2.9 | 3.5 | -0.6** |
| Low-use practice pattern | 3.0 | 2.5 | 0.5 |
| Subgroup difference | | | -1.1* |
| Transfer by One Year Postadmission, Less Restrictive Measure (Percentage) | | | |
| High-use practice pattern | 4.4 | 4.7 | -0.3 |
| Low-use practice pattern | 4.0 | 3.4 | 0.6 |
| Subgroup difference | | | -0.9 |

| | **High-Use Practice Pattern** | | **Low-Use Practice Pattern** | |
|---|---|---|---|---|
| **Sample Sizes** | **Selection** | **Retention** | **Selection** | **Retention** |
| **Treatment Group** | | | | |
| **Observations** | **16,154** | **14,218** | **15,964** | **12,166** |
| **Agencies** | **18** | **17** | **27** | **22** |
| **Control Group** | | | | |
| **Observations** | **10,162** | **9,689** | **12,073** | **16,060** |
| **Agencies** | **24** | **25** | **15** | **15** |

---

SOURCE: Selection measures obtained from Medicare claims, QA admission forms, UB-92 remarks fields, Medicare enrollment files, and agency cost reports. Retention measures obtained from Medicare claims data.

NOTE: Each observation in the sample reflects a patient's admission to (episode start in) a home health agency participating in the demonstration. The selection sample includes all available episodes. The retention sample includes only initial episodes in the first demonstration year. An agency is defined as having a high-use practice pattern if its (case-mix-adjusted) number of visits per episode in the base (predemonstration) period was above the median for all agencies in the sample. Otherwise, it is defined as having a low-use practice pattern.

[a] Measured at home health admission. Predicted utilization measures is based on the primary measure (which excludes planned treatments) and reflects service use during the first 120 days after admission (the at-risk period). Serious decubiti are diagnosed as stage 3 or 4.

[b] Base transfer measure: 0 to 3 days between discharge and other-agency admission, or 0 to 7 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date. Less restrictive measure: 0 to 7 days between discharge and other-agency admission, or 0 to 14 days between an other-agency claim that starts before the discharge date and a second claim that starts after the discharge date.

[c] Regression-adjusted to control for preexisting differences between treatment and control agencies.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

## TABLE D.5

## IMPACT OF PER-EPISODE PROSPECTIVE PAYMENT ON KEY PATIENT SELECTION AND RETENTION MEASURES, BY WHETHER THE AGENCY IS A CHAIN MEMBER

| | Treatment Group Mean[c] | Control Group Mean | Difference[c] (*p*-value)[d] |
|---|---|---|---|
| **Selection Measures** | | | |
| Predicted Service Utilization | | | |
| Chain member | 3,670 | 3,505 | 165 |
| Not chain member | 3,204 | 3,122 | 82 |
| Subgroup difference | | | 83 |
| Has Diabetes | | | |
| Chain member | 25.7 | 21.3 | 4.4** |
| Not chain member | 19.3 | 19.8 | -0.5 |
| Subgroup difference | | | 4.9** |
| Has Serious Decubiti Condition | | | |
| Chain member | 5.0 | 3.9 | 1.1 |
| Not chain member | 3.1 | 3.4 | -0.3 |
| Subgroup difference | | | 1.4 |
| Has Paid Help Available | | | |
| Chain member | 19.3 | 19.7 | -0.4 |
| Not chain member | 15.7 | 14.6 | 1.1 |
| Subgroup difference | | | -1.5 |
| **Retention Measures** | | | |
| Transfer by One Year Postadmission, Base Measure (Percentage) | | | |
| Chain member | 4.3 | 3.9 | 0.4 |
| Not chain member | 2.2 | 2.9 | -0.7* |
| Subgroup difference | | | 1.1 |
| Transfer by One Year Postadmission, Less Restrictive Measure (Percentage) | | | |
| Chain member | 5.6 | 4.8 | 0.8 |
| Not chain member | 3.2 | 3.9 | -0.7 |
| Subgroup difference | | | 1.5 |

| | **Chain Member** | | **Not Chain Member** | |
|---|---|---|---|---|
| **Sample Sizes** | **Selection** | **Retention** | **Selection** | **Retention** |
| **Treatment Group** | | | | |
| **Observations** | **9,565** | **9,460** | **22,333** | **16,984** |
| **Agencies** | **17** | **17** | **28** | **22** |
| **Control Group** | | | | |
| **Observations** | **3,281** | **3,717** | **18,954** | **22,032** |
| **Agencies** | **10** | **11** | **29** | **29** |

TABLE D.5 *(continued)*

---

SOURCE: Selection measures obtained from Medicare claims, QA admission forms, UB-92 remarks fields, Medicare enrollment files, and agency cost reports. Retention measures obtained from Medicare claims data.

NOTE: Each observation in the sample reflects a patient's admission to (episode start in) a home health agency. The selection sample includes all available episodes. The retention sample includes only initial episodes in the first demonstration year.

[a] Measured at home health admission. Predicted utilization measures is based on the primary measure (which excludes planned treatments) and reflects service use during the first 120 days after admission (the at-risk period). Serious decubiti (bed sores) are diagnosed as stage 3 or 4.

[b] Base transfer measure: 0 to 3 days between discharge and other-agency admission, or 0 to7 days between an other-agency claim that overlaps the discharge date and a second claim after the discharge date. Less restrictive measure: 0 to 7 days between discharge and other-agency admission, or 0 to 14 days between an other-agency claim that overlaps the discharge date and a second claim after the discharge date.

[c] Regression-adjusted to control for preexisting differences between treatment and control agencies.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

CMS Library
C2-07-13
7500 Security Blvd
Baltimore, MD 21244

CMS LIBRARY
3 8095 00005721 2